

NBER WORKING PAPER SERIES

SPARSE SIGNALS IN THE CROSS-SECTION OF RETURNS

Alexander M. Chinco
Adam D. Clark-Joseph
Mao Ye

Working Paper 23933
<http://www.nber.org/papers/w23933>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2017

We have received many helpful comments and suggestions from John Campbell, Victor DeMiguel, Xavier Gabaix, Andrew Karolyi, Bryan Kelly, Maureen O'Hara, Vassilis Papavassiliou, Ioanid Rosu, Thomas Ruchti, Gideon Saar, Allan Timmermann, Heather Tookes, Sunil Wahal, and Brian Weller as well as from seminar participants at the University of Illinois Urbana-Champaign, the 11th Annual Central Bank Conference on the Microstructure of Financial Markets, the 2016 AFA Annual Meetings, and the NBER EFFE SI. Hao Xu, Ruixuan Zhou, and Rukai Lou provided excellent research assistance. This research is supported by National Science Foundation grant #1352936, which is joint with the Office of Financial Research at the U.S. Department of the Treasury. This work also uses the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant #OCI-1053575. We thank David O'Neal of the Pittsburgh Supercomputer Center for his assistance with supercomputing, which was made possible through the XSEDE Extended Collaborative Support Service (ECSS) program. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Alexander M. Chinco, Adam D. Clark-Joseph, and Mao Ye. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Sparse Signals in the Cross-Section of Returns
Alexander M. Chinco, Adam D. Clark-Joseph, and Mao Ye
NBER Working Paper No. 23933
October 2017
JEL No. C55,C58,G12,G14

ABSTRACT

This paper applies the Least Absolute Shrinkage and Selection Operator (LASSO) to make rolling 1-minute-ahead return forecasts using the entire cross section of lagged returns as candidate predictors. The LASSO increases both out-of-sample fit and forecast-implied Sharpe ratios. And, this out-of-sample success comes from identifying predictors that are unexpected, short-lived, and sparse. Although the LASSO uses a statistical rule rather than economic intuition to identify predictors, the predictors it identifies are nevertheless associated with economically meaningful events: the LASSO tends to identify as predictors stocks with news about fundamentals.

Alexander M. Chinco
University of Illinois at Urbana-Champaign
1206 S 6th St, Rm 343J
Champaign, IL 61820
AlexChinco@gmail.com

Adam D. Clark-Joseph
University of Illinois at Urbana-Champaign
Champaign, IL
adcj@illinois.edu

Mao Ye
University of Illinois at Urbana-Champaign
343K Wohlers Hall
1206 South Sixth Street
Champaign, IL 61820
and NBER
maoye@illinois.edu

1 Introduction

Financial economists have been looking for variables that predict future stock returns for as long as there have been financial economists. For example, [Banz \(1981\)](#) used market cap to predict future returns, [Jegadeesh and Titman \(1993\)](#) used lagged returns, and [Cohen and Frazzini \(2008\)](#) used customer earnings surprises. To find these sorts of variables, researchers have to solve two distinct problems: identification and estimation. First, they have to identify a subset of candidate predictors; then, they have to estimate the quality of these predictors.

In the past, researchers typically reached for a different set of tools when working on each of these problems, using their intuition to identify predictors and statistics to estimate quality. This two-pronged approach works well when you are only looking for steady long-lived predictors. For example, Cabot Oil & Gas's lagged return predicted its future return at the 1-minute horizon throughout October 2010. Because it formed such a steady long-lived predictive relationship, a researcher could intuit this variable and then estimate its quality with an ordinary least-squares (OLS) regression:

$$r_{t-\ell} = \hat{\alpha} + \hat{\beta} \cdot x_{t-(\ell+1)} + \varepsilon_{t-\ell} \quad \ell \in \{0, \dots, L-1\} \quad (1)$$

Above, r_t is Cabot's minute- t return, $\hat{\alpha}$ is its mean return, L is the length of the estimation window, x_{t-1} is Cabot's lagged return standardized to have mean zero and unit variance during the estimation window, and $\hat{\beta}$ is the associated OLS coefficient.

But, modern financial markets are big, fast, and complex. Predictability now exists at scales that are not easy for a researcher to intuit. For instance, the lagged returns of Family Dollar Corp. were a significant predictor for 20% of the oil-and-gas industry—including Cabot Oil & Gas—during a 50-minute stretch on October 6th, 2010. Can a researcher really fish this particular variable out of the sea of spurious predictors using only his intuition? Of course not.

And, without a clear idea of which candidate predictors to test, a researcher cannot use the OLS regression in Equation (1) to estimate the amount of cross-stock predictability. There were 2,191 NYSE-listed stocks in October 2010. So, using an OLS regression to estimate the relationship between Cabot's current return and the lagged returns of every one of these candidate predictors would require at least 2,191 observations—or, nearly 6 trading days! A researcher cannot wait 6 days to identify a signal that lasts less than an hour.

Motivated by this problem, we apply the Least Absolute Shrinkage and Selection

Operator (LASSO) rather than intuition to identify unexpected short-lived predictors, such as the lagged returns of Family Dollar Corp. We find that using the LASSO increases both out-of-sample fit and forecast-implied Sharpe ratios, and we show that this out-of-sample success comes from identifying predictors that are unexpected, short-lived, and sparse. Finally, we document that these predictors are often the lagged returns of stocks with news about fundamentals. In other words, although the unexpected short-lived predictors that the LASSO identifies are not easy to intuit, they are still economically meaningful.

The LASSO. We begin our analysis by describing both how the LASSO works and why we use it. We are motivated by a simple observation. As researchers, we cannot use our intuition to identify predictors that are sufficiently unexpected and short-lived—at some point, our brains just do not work fast enough. So, to incorporate these sorts of signals into our return forecasts, we need some other way to solve the identification problem described above.

One popular approach is to assume that there are only a handful of important predictors at any one point in time—that is, to bet on sparsity. Morally speaking, if only $S \ll 2,191$ predictors are important for forecasting Cabot’s returns, then you should only need a few more than S observations to identify and estimate this sparse set of predictors. Under this assumption, a researcher trying to forecast the returns of Cabot Oil & Gas can use the LASSO to incorporate unexpected short-lived signals, such as the lagged return of Family Dollar Corp.

Using the LASSO means solving the following optimization problem:

$$\tilde{\alpha}, \tilde{\beta} = \arg \min_{\alpha, \beta \in \mathbf{R}^{2,192}} \left\{ \frac{1}{L} \cdot \sum_{\ell=0}^{L-1} \left(r_{t-\ell} - \alpha - \sum_{n'=1}^{2,191} \beta_{n'} \cdot x_{n',t-(\ell+1)} \right)^2 + \lambda \cdot \sum_{n'=1}^{2,191} |\beta_{n'}| \right\} \quad (2)$$

Just like in Equation (1), L is still the length of the estimation window, r_t is still Cabot’s minute- t return, and $\tilde{\alpha}$ is still its mean return. But, there are two differences. Now, the optimization problem includes the lagged returns of all $N > 1$ NYSE-listed stocks instead of just Cabot’s lagged return. And, there is a new penalty function, $\lambda \cdot \sum_{n'} |\beta_{n'}|$. Note that we will use tildes rather than hats to distinguish LASSO coefficient estimates from the corresponding OLS coefficient estimates.

Although this equation looks complicated at first, it is not. It is a simple extension of the OLS regression in Equation (1). In fact, if you were to ignore the right-most term—the penalty function, $\lambda \cdot \sum_{n'} |\beta_{n'}|$ —then it would simply be an OLS regression.

The Univariate LASSO

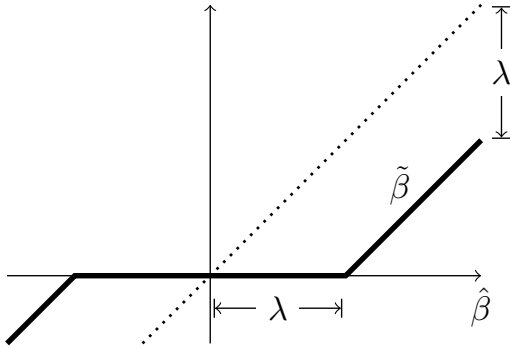


Figure 1: *x-axis: OLS estimate, $\hat{\beta}$. Solid line: LASSO estimate, $\tilde{\beta}$, given the penalty parameter, λ . Dotted line: $y = x$. (Reads) “If the OLS estimate is more extreme than the LASSO’s penalty parameter, $|\hat{\beta}| > \lambda$, then the LASSO estimate will be a shrunken version of the OLS estimate, $|\tilde{\beta}| = |\hat{\beta}| - \lambda > 0$. But, if the OLS estimate is less extreme than the LASSO’s penalty parameter, $|\hat{\beta}| \leq \lambda$, then the LASSO estimate will be precisely zero, $\tilde{\beta} = 0$.”*

But, it is this penalty function that allows the LASSO to exploit a bet on sparsity.

To see why, consider the solution to Equation (2) when there is only $N = 1$ predictor, like in Equation (1), instead of $N = 2,191$:

$$\tilde{\beta} = \begin{cases} \text{Sign}[\hat{\beta}] \cdot (|\hat{\beta}| - \lambda) & \text{if } |\hat{\beta}| > \lambda \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Above, $\text{Sign}[z] = z/|z|$ and $\hat{\beta}$ represents the OLS coefficient computed in Equation (1). Equation (3) shows that, if the OLS coefficient is more extreme than the LASSO’s penalty parameter, $|\hat{\beta}| > \lambda$, then the LASSO will estimate a shrunken version of the OLS coefficient, $|\tilde{\beta}| = |\hat{\beta}| - \lambda > 0$. Whereas, if the OLS coefficient is less extreme than the LASSO’s penalty parameter, $|\hat{\beta}| \leq \lambda$, then the LASSO will estimate a value of precisely zero, $\tilde{\beta} = 0$. Figure 1 depicts this relationship. Thus, if the unexpected short-lived predictors are sparse—that is, if only $S \leq L$ of these predictors are more extreme than λ —then the LASSO will only have to estimate the $S \leq L$ coefficients larger than λ . And, as a result, a researcher can use the LASSO to incorporate these signals into his return forecast using a few more than S observations.

Out-of-Sample Performance. After describing how and why we use the LASSO to bet on sparsity, we next investigate whether this bet pays off. To do this, we evaluate 1-minute-ahead return forecasts for a randomly chosen subset of 250 NYSE-listed stocks on each trading day from January 2005 to December 2012. As a benchmark, we start with forecasts made via OLS regressions that include only steady long-lived predictors, such as a stock’s own lagged returns or the lagged returns on the market. In our main specifications, we study 1-minute-ahead return forecasts created using 3 lags of various steady long-lived predictors, but the exact number of lags does not

qualitatively affect our results.

We then apply the LASSO to make these same 1-minute-ahead return forecasts using the lagged returns of all 2,000+ NYSE-listed stocks during the previous 3 minutes as candidate predictors. We find that using the LASSO in addition to a standard benchmark model increases out-of-sample fit by at least $\Delta \bar{R}_n^2 = 1.2\%$ pt relative to just using the benchmark model by itself. A 1.2%pt increase might seem small, but remember that we are making 1-minute-ahead return forecasts. And, at short horizons, small increases in “ R^2 statistics can generate large benefits for investors (Campbell and Thompson, 2008, p. 1526)”. To highlight this point, we convert the LASSO’s 1-minute-ahead return forecasts into a forecast-implied trading strategy and document that this forecast-implied strategy generates an annualized Sharpe ratio of 1.8 *net* of trading costs.

At this point, it is important to emphasize two things about these results. The first is that we are running out-of-sample tests. It should not be surprising that the LASSO has better in-sample fit than an OLS regression. The LASSO can choose from over $3 \cdot N \approx 6,000$ candidate predictors; in a 30-minute estimation window, an OLS regression is restricted to fewer than 30. But, there is no guarantee that the LASSO’s better in-sample fit will translate into better out-of-sample fit. This will only happen if the cross-section of returns actually contains a sparse collection of $S < 30$ signals. If there are no signals to be found or if there are more than 30 signals, then using the LASSO will not help.

The second is that our results do not imply that researchers should use the LASSO *instead of* the standard two-pronged approach—that is, instead of their intuition. As a researcher, if you can use your economic intuition to identify a steady long-lived predictor, such as a stock’s own lagged returns, then an OLS regression is the right way to incorporate this predictor into your return forecast (Abadie and Kasy, 2017). But, the Bible does not say that all sources of return predictability should make intuitive sense.¹ And, the fact that using the LASSO and a benchmark model increases out-of-sample fit relative to just using the benchmark model suggests that researchers are ignoring important sources of predictability at the 1-minute horizon when they restrict their attention to steady long-lived predictors that they can suss out using nothing more than their intuition.

¹Original quote by Enrico Fermi: “The Bible does not say that all laws of nature are expressible linearly.” Found in: Ulam, S (1976). *Adventures of a Mathematician*. New York: Scribners.

Predictor Analysis. After seeing that the LASSO increases out-of-sample fit, a natural next question is: What do the predictors selected by the LASSO look like? Consistent with our original motivation, we find that the predictors selected by the LASSO are unexpected, short-lived, and sparse.

We begin by looking at why the predictors are unexpected. When we examine the LASSO's cross-validated penalty parameter, we find that it has an average value of $\lambda = 2.5\%$ per month. In other words, the LASSO ignores any predictor weaker than 2.5% per month when making its 1-minute-ahead return forecasts. This lower bound is twice as large as well-known predictors at the weekly and monthly horizon documented in the academic literature (McLean and Pontiff, 2016; Harvey et al., 2016; Linnainmaa and Roberts, 2016). As a result, there is little correlation between the predictors identified by the LASSO each minute and existing predictors documented in the academic literature, which makes the LASSO's choice of predictors seem unexpected. In addition, the predictors selected by the LASSO are also short-lived and sparse. Less than 5% of the predictors selected by the LASSO are used for more than 15 minutes in a row. And, on average the LASSO uses the lagged returns of only 12.7 other stocks as predictors when making its return forecast for each stock.

Economic Origins. We conclude our analysis by investigating the economic, rather than statistical, origins of the LASSO's success. This is important because the LASSO uses a purely statistical rule to identify candidate predictors that are too unexpected and short-lived to be pinned down by economic intuition alone. And, because it does not rely on a researcher's intuition, the LASSO's increase in out-of-sample fit could in principle have nothing to do with economic fundamentals. For example, it could be the case that the LASSO's increase in out-of-sample fit comes from identifying statistical artifacts. But, this is not what we find in the data. We document that the LASSO tends to identify as predictors the lagged returns of stocks with news about fundamentals. In other words, although the unexpected short-lived predictors that the LASSO identifies are not easy to intuit, they are still economically meaningful.

1.1 Literature

This paper builds on several strands of the asset-pricing and statistics literature.

Predictability. To start with, our paper relates to work on return predictability (McLean and Pontiff, 2016; Harvey et al., 2016; Linnainmaa and Roberts, 2016).

Campbell and Thompson (2008) shows that many of the predictors documented in the academic literature work much better in sample. And, DeMiguel et al. (2009) documents that simpler predictive regressions perform better out of sample due to overfitting. The LASSO avoids this overfitting problem by using a penalty function that removes all but the strongest predictors.

Sparsity. Our paper also relates to work on the role of sparsity in financial economics. For example, Gabaix (2014) proposes that traders prune away the least important predictors when forming mental models by internally imposing an ℓ_1 penalty. Chinco (2015) shows that, if traders have to uncover sparse signals in market data, then there are information-theoretic limits to how quickly they can interpret what the market is telling them.

The LASSO. In addition, this paper builds on a large body of work looking at the performance of penalized regressions, such as the LASSO. The LASSO was first introduced in Tibshirani (1996). See Hastie et al. (2001) for a general introduction. Meinshausen and Yu (2009) shows how LASSO-type estimators extend to settings with correlated right-hand-side variables.

There are several excellent papers applying these ideas to topics in financial economics. For example, DeMiguel et al. (2009) explains why norm-constrained estimators will perform better out of sample in the presence of estimation error. Both Freyberger et al. (2017) and Feng et al. (2017) use the LASSO to identify characteristics related to the cross-section of expected returns at the monthly horizon. Bryzgalova (2017) develops a new model-specification test by imposing an ℓ_1 -penalty function. And, Kozak et al. (2017) shows that using ℓ_1 - and ℓ_2 -penalized regressions to forecast returns can increase forecast-implied Sharpe ratios using monthly data.

Our paper differs from these other financial applications of the LASSO in an important way. These papers are using the LASSO to choose between various predictors that have already been identified by other researchers using the standard two-pronged approach. By contrast, we are studying a situation where the standard two-pronged approach simply does not apply. A researcher cannot use his intuition to identify candidate predictors that are sufficiently unexpected and short-lived. And, without a clear idea of which candidate predictors to test, a researcher cannot use an OLS regression to incorporate these cross-stock signals into his return forecasts. So, we are suggesting the LASSO should be used as a complement to existing methods—that is, a way for researchers to analyze unexpected short-lived predictors that they would

not otherwise have access to using existing methods.

News. Finally, our paper relates to the literature looking at the relationship between news and asset prices. Brogaard et al. (2014) show that algorithmic traders' orders predict price changes over very short horizons and are correlated with macroeconomic news announcements. And, Manela and Moreira (2017) uses news to form an uncertainty index and shows that it is related to expected returns. Theoretical models of algorithmic trading (e.g., Hoffman, 2014; Foucault et al., 2016; Du and Zhu, 2016) focus on this ability to react to news announcements faster for individual firms. By contrast, our results suggest that the LASSO success comes from quickly identifying the unexpected consequences of news announcements for *other* firms.

1.2 Notation

Throughout our analysis, Greek letters will denote in-sample parameter values. A superscript star* will denote a parameter value from the true data-generating process. A hat will denote an in-sample OLS estimate of the true parameter value during a particular 30-minute estimation window, and a tilde will denote the corresponding in-sample LASSO estimate. Although these in-sample estimates are functions of the particular 30-minute estimation window they were estimated in, we will not include an estimation-window subscript to avoid clutter. **Boldface characters** will denote vectors. Finally, we will use a bar to denote out-of-sample parameter estimates.

2 The LASSO

We now outline both how the LASSO works and why we use it.

2.1 Motivation

To see why we use the LASSO to make 1-minute-ahead return forecasts, consider the following data-generating process for the n th stock's returns in minute t .

Data-Generating Process. Suppose that the n th stock's return in minute t could be related to the lagged returns of any of the other N stocks in the market during

the previous 3 minutes:

$$r_{n,t} = \alpha_n^* + \sum_{n'=1}^{3 \cdot N} \beta_{n,n'}^* \cdot x_{n',t-1} + \varepsilon_{n,t}^* \quad (4)$$

In the equation above, α_n^* is the mean return of the n th stock, and $\varepsilon_{n,t}^*$ is a noise term with $E[\varepsilon_{n,t}^*] = 0$ and $\text{Var}[\varepsilon_{n,t}^*] = \sigma_\varepsilon^2$. Since we are considering as predictors the lagged return of each NYSE-listed stock during the previous 3 minutes, we now use $x_{n',t-1}$ to denote the value of one of these $3 \cdot N$ predictors standardized to have mean zero and unit variance during the estimation window. The parameter $\beta_{n,n'}^*$ is then the predictive power of the n' th predictor for the n th stock's returns. In our empirical analysis, we will be making 1-minute-ahead return forecasts for many different stocks, not just for Cabot Oil & Gas. So, from now on, we will include an n subscript to denote which stock we are interested in forecasting.

Unexpected and Short-Lived. If the predictive relationship embodied by $\beta_{n,n'}^*$ is steady and long-lived, then it is reasonable to think that a financial economist could intuit this connection. After all, researchers have used their intuition to uncover hundreds if not thousands of cross-sectional return predictors (McLean and Pontiff, 2016; Harvey et al., 2016; Linnainmaa and Roberts, 2016). And, if a researcher can intuit a candidate predictor, then he can use an OLS regression to estimate the quality of this predictor. This is the standard two-pronged approach to finding predictors mentioned in the introduction.

But, if a researcher cannot use his intuition to narrow down the list of candidate predictors, then an OLS regression is no longer a viable option. For example, to estimate all $3 \cdot N$ of the cross-stock predictive relationships in Equation (4) using an OLS regression, you would have to solve the optimization problem below

$$\hat{\alpha}_n, \hat{\beta}_n \stackrel{\text{def}}{=} \arg \min_{\alpha_n, \beta_n} \left\{ \frac{1}{L} \cdot \sum_{\ell=0}^{L-1} \left(r_{n,t-\ell} - \alpha_n - \sum_{n'=1}^{3 \cdot N} \beta_{n,n'} \cdot x_{n',t-(\ell+1)} \right)^2 \right\} \quad (5)$$

where $\hat{\beta}_n = [\hat{\beta}_{n,1} \ \cdots \ \hat{\beta}_{n,3 \cdot N}]^\top$ is the $(3 \cdot N \times 1)$ -dimensional vector of slope coefficients used to forecast the returns of the n th stock.

This multivariate analogue to Equation (1) is ill-posed when there are more parameters to estimate than observations to estimate them with, $(3 \cdot N + 1) > L$. There are roughly $N = 2,000$ NYSE-listed stocks at any point in time, and 2,000 minutes is roughly 6 trading days. So, if the cross-stock signal represented by $\beta_{n,n'}^*$ does not stick around for at least 18 trading days, then there is no way for a researcher to

estimate it using an OLS regression without making some sort of intuitive leap. If a researcher wants to use such an unexpected short-lived predictor, then he needs to find some other way of solving his identification problem which shortens the required estimation window.

Betting on Sparsity. One popular approach is to assume that there are only a handful of strong predictors at any point in time—that is, to bet on sparsity (Hastie et al., 2001). Morally speaking, it should not require all $3 \cdot N$ observations to estimate a coefficient vector containing mostly zeros. If $S_{n,\lambda}$ denotes the number of predictors for stock n that are more extreme than λ_n

$$S_{n,\lambda} \stackrel{\text{def}}{=} \sum_{n'=1}^{3 \cdot N} 1_{\{|\beta_{n,n'}^*| > \lambda_n\}} \quad (6)$$

then betting on sparsity with an L -minute estimation window means assuming that:

$$S_{n,\lambda} \in \{1, 2, \dots, L - 1\} \quad (7)$$

It is important to note that this bet is not guaranteed to succeed. There are two ways for the data to be non-sparse. First, if $S_{n,\lambda} = 0$, then there are no predictors to find. Second, if $S_{n,\lambda} \geq L$, then the predictors are dense, meaning that there is no way to consistently estimate the relationship between all these predictors and the n th stock's returns in an L -minute estimation window no matter what approach you take. Either way, if the predictors are not sparse, then a researcher's bet on sparsity will not pay off.

Why Sparsity Helps. Before getting to any more math, let's pause for a moment and build some intuition about why betting on sparsity might help. Suppose you are standing on a beach, and you are trying to figure out whether a dark spot you see on the horizon is a sailboat or a cloud. A natural human reaction in this situation is to squint. But, why are our brains wired this way? After all, by squinting and slightly closing your eyes, you are removing information and making it harder to see colors and quick movements.

The key insight is that colors and quick movements are not going to tell you much about the nature of the mysterious blob on the horizon. Sailboats and clouds are both going to be dark and slow moving at this distance. So, in essence, squinting penalizes these weak predictors. When you squint, the only things left for you to see are the most relevant details for the question you are trying to answer: Is the blob triangular? Does it have a mast? And, by ignoring extraneous details, you will be

able identify the nature of the object on the horizon at a much greater distance.

But, this approach will only work if you can answer your question using only a handful of strong predictors. For instance, squinting will do you no good when trying to figure out whether the sailboat is flying a British or American flag. Both of these flags use the same colors. Distinguishing between these two possibilities will require many subtle cues about how the colors are organized. No amount of squinting will help you resolve these subtle differences. You will just have to wait until the sailboat gets closer and get a better look.

Subset Selection. Most economists are familiar with the Akaike information criterion (AIC; Akaike, 1974) and the Bayesian information criterion (BIC; Schwarz, 1978). These criteria are examples of a particular way of solving the selection problem known as best-subset selection. They both impose a penalty that eliminates weak predictors—just like in the example above—by solving the following optimization problem for different choices of λ_n :

$$\min_{\alpha_n, \beta_n} \left\{ \frac{1}{L} \cdot \sum_{\ell=0}^{L-1} \left(r_{n,t-\ell} - \alpha_n - \sum_{n'=1}^{3 \cdot N} \beta_{n,n'} \cdot x_{n',t-(\ell+1)} \right)^2 + \lambda_n \cdot \sum_{n'=1}^{3 \cdot N} 1_{\{\beta_{n,n'} \neq 0\}} \right\} \quad (8)$$

If $\lambda_n = 2 \cdot \sigma_\varepsilon$, then solving this problem above is equivalent to using the AIC; whereas, if $\lambda_n = 2 \cdot \sigma_\varepsilon \cdot \log(L)$, then it is equivalent to using the BIC (Foster and George, 1994).

So, why not use the AIC or the BIC to identify unexpected short-lived predictors in the cross-section of returns? The problem with these criteria is that they “are highly impractical (Candes and Plan, 2009, p. 2146)” when choosing among many variables. When solving the optimization problem in Equation (8), there is a fixed λ_n penalty for moving $\beta_{n,n'}$ any distance away from zero, no matter whether it is an inch or a mile. And, this hard threshold implies that the optimization problem in Equation (8) is non-convex and thus NP hard (Natarajan, 1995), requiring an exhaustive search over all possible subsets of $2^{3 \cdot N}$ predictors.

To give a sense of the difficulty of this combinatorial-search problem, note that WiFi networks are considered secure because cracking a 128-bit passcode would involve a brute-force search over 2^{128} parameter values. Such a search would take 1 billion billion years (Arora, 2012). So, running more than $2^{6,000}$ regressions is out of the question.

Convex Relaxation. The difficulty of this combinatorial-search problem is what brings us to the LASSO. The LASSO, whose name is an acronym for least absolute

Estimation and Forecast Timing

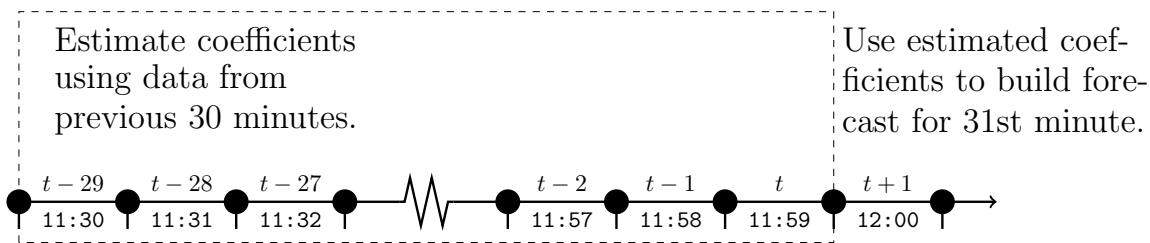


Figure 2: To make our 1-minute-ahead forecast for the n th stock’s return in minute $(t + 1) = 12:00$, we first estimate a model using data from the previous 30 minutes, $\{11:30, \dots, 11:59\}$. Then, we apply the estimated coefficients to the most recent 3-minutes of data, $\{11:57, 11:58, 11:59\}$. We use $f_{11:59}^{\text{LASSO}}$ to denote this forecast for the n th stock’s return in minute 12:00, because it only uses information up to minute 11:59.

shrinkage and *selection* operator, offers a way to solve this subset-selection problem that gets around the computational roadblock associated with AIC and BIC. The LASSO does this by changing the penalty function in Equation (8) slightly:

$$\tilde{\alpha}_n, \tilde{\beta}_n \stackrel{\text{def}}{=} \arg \min_{\alpha_n, \beta_n} \left\{ \frac{1}{L} \cdot \sum_{\ell=0}^{L-1} \left(r_{n,t-\ell} - \alpha_n - \sum_{n'=1}^{3 \cdot N} \beta_{n,n'} \cdot x_{n',t-(\ell+1)} \right)^2 + \lambda_n \cdot \sum_{n'=1}^{3 \cdot N} |\beta_{n,n'}| \right\} \quad (9)$$

The optimization problem in Equation (8) uses the penalty $1_{\{\beta_{n,n'} \neq 0\}}$ while the problem in Equation (9) uses the penalty $|\beta_{n,n'}|$. This is the only difference between the two optimization problems. But, this difference convexifies the problem (Efron and Hastie, 2016, p. 308). And, “if we can formulate a problem as a convex optimization problem, then we can solve it efficiently. . . with only a bit of exaggeration, we can say that, if you formulate a practical problem as a convex optimization problem, then you have solved the original problem (Boyd and Vandenberghe, 2004, p. 9).”

2.2 Implementation

With this motivation in mind, let’s now examine the details of how we use the LASSO to make 1-minute-ahead return forecasts in rolling 30-minute estimation windows using the entire cross-section of lagged returns as candidate predictors.

LASSO Forecasts. Prior to the start of each trading day, we choose 250 stocks at random. Then, for each stock n in this collection of 250, we solve the optimization problem in Equation (9) in R using the `glmnet` package (Friedman et al., 2010) in rolling ($L = 30$)-minute estimation windows. As predictors, we consider the lagged

returns of all NYSE-listed stocks during the previous 3 minutes, including the lagged returns of stock n itself. This means solving an estimation problem with $3 \cdot N \approx 6,000$ right-hand-side variables using only 30 observations, something that would clearly be impossible within an OLS-regression framework. We generate a 1-minute-ahead return forecast, $f_{n,t}^{\text{LASSO}}$, for the n th stock’s return in minute $(t + 1)$ by applying the LASSO-coefficient estimates from the previous 30 minutes to the most recent 3 minutes of data:

$$f_{n,t}^{\text{LASSO}} \stackrel{\text{def}}{=} \tilde{\alpha}_n + \sum_{n'=1}^{3 \cdot N} \tilde{\beta}_{n,n'} \cdot x_{n',t} \quad (10)$$

Figure 2 summarizes this procedure for creating 1-minute-ahead return forecasts.

Penalty Parameter. There is a clear choice for λ_n in best-subset-selection procedures like the AIC and the BIC. The LASSO convexifies these problems, making them computationally tractable in settings with thousands of predictors. But, outside of special cases, there is no theoretically optimal value for λ_n when using the LASSO. So, we follow the standard procedure (Hastie et al., 2001) and choose λ_n on a stock-by-stock basis in each 30-minute estimation window via K -fold cross-validation.

Alternative Penalties. The LASSO is not the only way to convexify the best-subset-selection problem in Equation (8). Researchers have proposed alternative procedures, such as the adaptive LASSO (Zou, 2006) or the elastic net (Zou and Hastie, 2005). These procedures represent interesting variations on the LASSO, but we focus on the LASSO for a simple reason: we want to show that it is possible for researchers to incorporate unexpected short-lived signals into their existing return forecasts using a statistical selection rule rather than their intuition. To accomplish this goal, it is best to use the simplest possible version of one of these rules—namely, the LASSO.

Benchmark Forecasts. Finally, in the section below, we show that using 1-minute-ahead return forecasts from both the LASSO and a standard benchmark model increases out-of-sample fit relative to just using the return forecast from the benchmark model. To construct these benchmark forecasts, we run OLS regressions that include a wide range of steady long-lived predictors. We estimate each of these regressions in the exact same rolling 30-minute estimation windows described above. We then compute our benchmark 1-minute-ahead return forecasts, $f_{n,t}^{\text{Bmk}}$, by applying the resulting OLS estimates to the most recent 3 minutes of data. In our main specifications, we use benchmark models with 3 lags of each predictor, but we show that the exact

number of lags does not qualitatively affect the results.

3 Out-of-Sample Performance

This section investigates whether it pays to bet on sparsity with the LASSO. This means testing whether using both the LASSO and a benchmark model to forecast returns increases out-of-sample fit relative to just using the benchmark model.

3.1 Data Description

Our tests incorporate data from several different sources.

1-Minute Returns. We get data on the 1-minute returns of NYSE-listed stocks from the Trade and Quote (TAQ) database. Our sample includes every trading day from January 2005 to December 2012. We remove stocks with prices below \$5 at the previous day's close. While we consider the entire cross-section of NYSE-listed stocks as candidate predictors, we only make 1-minute-ahead return forecasts for a randomly selected subset of 250 stocks each day for computational reasons.

Because we use a 30-minute estimation window and 3 lags, the first 1-minute-ahead return forecast each day is at $t = 10:04\text{am}$. To avoid distortions due to the NYSE closing auction, we also drop the last minute of each trading day from our sample. This means that we make 356 separate 1-minute-ahead return forecasts, $t \in \{10:04\text{am}, \dots, 3:59\text{pm}\}$, for each of the randomly selected 250 stocks during each day. In other words, each stock-day contains $356 \cdot 250 = 64,000$ forecasts.

Benchmark Predictors. We consider multiple steady long-lived predictors when making our benchmark 1-minute-ahead return forecasts. In autoregressive specifications, a stock's lagged returns come from the TAQ data described above. When a specification involves the lagged returns on a portfolio (market, size, or value), we use the 1-minute returns of exchange-traded funds (ETFs). The market return corresponds to the return on the iShares market ETF. And, size and value returns correspond to the returns on the iShares Russell 1000 and 2000 value and growth ETFs. When a specification involves lagged industry returns, we construct a value-weighted return for the corresponding 3-digit SIC industry.

Control Variables. We look at how the LASSO's increase in out-of-sample fit varies across stocks with different firm characteristics using data from CRSP. Market

capitalization is defined at the close of the previous trading day. Trading volume is defined as the total trading volume on the previous trading day. Return volatility is defined as the volatility of 1-minute returns during the previous trading day. And, bid-ask spread is the average bid-ask spread during the previous trading day.

Daily Factors. Finally, to compute the abnormal returns, we use data on the daily excess returns of factor portfolios from Kenneth French’s website.²

3.2 Out-of-Sample Fit

Before thinking about increases in out-of-sample fit, let’s start by asking: How much return variation can the LASSO explain? This estimate will provide an upper bound on the LASSO’s increase in out-of-sample fit because some of the return variation explained by the LASSO might already be explained by an existing benchmark model.

Econometric Approach. To answer this question, we estimate a suite of predictive regressions:

$$r_{n,t+1} = \bar{a}_n + \bar{b}_n \cdot \left(\frac{f_{n,t}^{\text{LASSO}} - \bar{m}_n^{\text{LASSO}}}{\bar{s}_n^{\text{LASSO}}} \right) + e_{n,t+1} \quad (11)$$

In the equation above, $r_{n,t+1}$ is the n th stock’s realized return in minute $(t+1)$, $f_{n,t}^{\text{LASSO}}$ is the LASSO’s 1-minute-ahead return forecast for minute $(t+1)$, \bar{m}_n^{LASSO} and \bar{s}_n^{LASSO} denote the mean and standard deviation of the LASSO’s 1-minute-ahead return forecasts for the n th stock on a given day, \bar{a}_n and \bar{b}_n denote out-of-sample regression coefficients, and $e_{n,t+1}$ is the forecasting error. We estimate a separate regression for each of the 250 randomly selected stocks on a given trading day, which means that each regression contains 356 observations. And, we normalize the LASSO’s 1-minute-ahead return forecasts to have mean zero and unit variance so that we can compare the slope coefficients, \bar{b}_n , across different regressions.

For each predictive regression, we then measure the fraction of the variation in a stock’s 1-minute returns on a given day that is explained by the LASSO using the adjusted R^2 statistic. At one extreme, an adjusted $\bar{R}_n^2 = 100\%$ means that the LASSO’s 1-minute-ahead return forecast, $f_{n,t}^{\text{LASSO}}$, explains all of the variation in a stock’s returns. Whereas, at the other extreme, an adjusted $\bar{R}_n^2 = 0\%$ means that $f_{n,t}^{\text{LASSO}}$ explains none of this variation.

Empirical Results. Table 1 summarizes the results of the suite of predictive re-

²http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Out-of-Sample Fit, The LASSO

	Mean	95% CI
\bar{a}_n [%/m]	0.002 (0.002)	.
\bar{b}_n [%/m]	1.433 (0.017)	[1.399, 1.467]
\bar{R}_n^2 [%]	2.467 (0.027)	[2.414, 2.520]

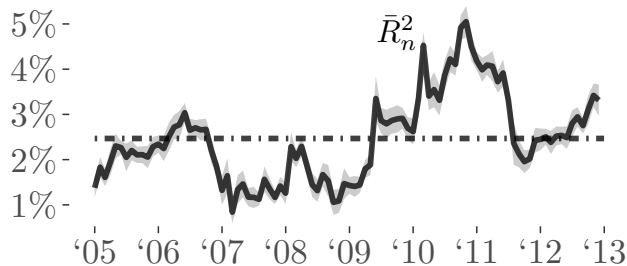


Table 1: *The LASSO’s out-of-sample fit. (Sample) Results of predictive regressions run at the stock-day level for 250 randomly chosen stocks on each trading day from January 2005 to December 2012. (Left) Population averages for regression coefficients and adjusted R^2 statistic. Numbers in parentheses are standard errors clustered by stock-day. 95% CI reports 95%-confidence intervals for population averages. (Right) Average adjusted R^2 statistic each month with the dashed line representing $\bar{R}_n^2 = 2.467\%$. Grey bands denote the 99.9% confidence interval computed using standard errors clustered by stock-day. (Reads) “On average, the LASSO’s 1-minute-ahead return forecast explains 2.467% of the variation in a stock’s returns on a given day. And, the LASSO explains at least 1% of the variation in returns during every month in our sample.”*

gressions that we run. On average, the LASSO’s 1-minute-ahead return forecast explains 2.467% of the variation in a stock’s 1-minute returns on a given day. The 95%-confidence interval shows that this point estimate is statistically different from zero. We two-way cluster the standard errors at the stock-day level to account for the possibility that the LASSO might do a better job of forecasting returns for particular stocks or on particular trading days. The figure on the right of Table 1 reveals that, while there is some time-series variation in the LASSO’s out-of-sample fit, the LASSO explains at least 1% of the variation in returns during every month in our sample.

3.3 Increase in Out-of-Sample Fit

Of course, we are not directly interested in the LASSO’s out-of-sample fit in an absolute sense. Our goal is to show that using the LASSO in addition to a benchmark model increases out-of-sample fit relative to just using the benchmark model. We are interested in increases in out-of-sample fit not the level of out-of-sample fit itself.

Illustrative Example. Consider an example showing why this distinction is important. Recall that the LASSO can identify a stock’s own lagged returns as candidate

predictors. So, in principle, the LASSO could be doing nothing more than replicating the 1-minute-ahead return forecast of an autoregressive model with 3 lags. If this were the case, then the LASSO would have significant out-of-sample fit in an absolute sense. But, it would not provide any additional information over and above what was already contained in the AR(3) model. To justify using the LASSO, we need to show that the 2.467% of the return variation explained by the LASSO is not already explained by an existing benchmark model. In other words, we need to show that using the LASSO increases out-of-sample fit.

Econometric Approach. We do this by elaborating on the methodology described above. First, we estimate the out-of-sample fit of a benchmark model just like we did for the LASSO:

$$r_{n,t+1} = \bar{a}_n + \bar{c}_n \cdot \left(\frac{f_{n,t}^{\text{Bmk}} - \bar{m}_n^{\text{Bmk}}}{\bar{s}_n^{\text{Bmk}}} \right) + e_{n,t+1} \quad (12)$$

In the specification above, $r_{n,t+1}$ is the n th stock's realized return in minute $(t + 1)$, $f_{n,t}^{\text{Bmk}}$ is the 1-minute-ahead return forecast for this stock's minute- $(t + 1)$ return generated by some benchmark model, \bar{m}_n^{Bmk} and \bar{s}_n^{Bmk} denote the mean and standard deviation of the benchmark model's 1-minute-ahead return forecasts, \bar{a}_n and \bar{c}_n are out-of-sample regression coefficients, and $e_{n,t+1}$ is the benchmark forecasting error.

Then, to see if the LASSO increases out-of-sample fit relative to this benchmark model, we estimate a second specification that includes both $f_{n,t}^{\text{LASSO}}$ and $f_{n,t}^{\text{Bmk}}$:

$$r_{n,t+1} = \bar{a}_n + \bar{b}_n \cdot \left(\frac{f_{n,t}^{\text{LASSO}} - \bar{m}_n^{\text{LASSO}}}{\bar{s}_n^{\text{LASSO}}} \right) + \bar{c}_n \cdot \left(\frac{f_{n,t}^{\text{Bmk}} - \bar{m}_n^{\text{Bmk}}}{\bar{s}_n^{\text{Bmk}}} \right) + e_{n,t+1} \quad (13)$$

If including the LASSO's 1-minute-ahead return forecast increases the adjusted R^2 statistic, $\Delta \bar{R}_n^2 \stackrel{\text{def}}{=} \bar{R}_n^{2,\text{Both}} - \bar{R}_n^{2,\text{Bmk}}$, then we know that the LASSO must be capturing new information that is not already present in the benchmark model's 1-minute-ahead return forecast.

At one extreme, if the LASSO and the benchmark model are using totally different information sets to create their respective 1-minute-ahead return forecasts, then we will estimate $\Delta \bar{R}_n^2 = 2.467\% \text{pt}$. At the other extreme, if the LASSO is just replicating the 1-minute-ahead return forecast of the benchmark model, then we will estimate $\Delta \bar{R}_n^2 = 0\% \text{pt}$. We are working with increases not levels. And, because we are studying percentage-point increases in adjusted R^2 , we can use the analysis in [Giacomini and White \(2006\)](#) to evaluate statistical significance (see Appendix B).

Empirical Results. Table 2 summarizes the LASSO's increase in out-of-sample

fit relative to a variety of different benchmark models. The first row reveals that 1-minute-ahead return forecasts made using an AR(3) model explain $\bar{R}_n^{2,\text{Bmk}} = 7.365\%$ of return variation on average. The out-of-sample fit of this benchmark model is larger than the LASSO’s out-of-sample fit. But, we are not suggesting that researchers should use the LASSO instead of their preferred benchmark model. We are suggesting that researchers should use the LASSO *in addition* to their preferred benchmark model. And, by using the LASSO in addition to this benchmark model, we estimate that a researcher could explain an additional $\Delta\bar{R}_n^2 = 1.185\%$ pt of the variation in returns. The [Giacomini and White \(2006\)](#) p -value shows that such a large increase in out-of-sample fit is highly unlikely under the null hypothesis that $\Delta\bar{R}_n^2 = 0\%$ pt.

The second row of Table 2 performs the exact same exercise, only this time relative to a benchmark model using 3 lags of the market return. This market benchmark performs much worse than the AR(3) benchmark, explaining only $\bar{R}_n^{2,\text{Bmk}} = 0.311\%$ of the variation in returns. And, the 1-minute-ahead return forecasts coming from the market benchmark use totally different information than those coming from the LASSO. The point estimate of $\Delta\bar{R}_n^2 = 2.469\%$ pt is identical to our original estimate for $\bar{R}_n^{2,\text{LASSO}} = 2.467\%$ to within 1 part in 500. What’s more, the third row shows that a combined benchmark model that uses a stock’s own returns as well as the market’s returns does not improve on the simple AR(3) benchmark.

Alternative Lag Structures. The next four rows in Table 2 examine the LASSO’s increase in out-of-sample fit relative to alternative AR(1), AR(2), AR(4), and AR(5) benchmark models. The AR(3) benchmark has the highest out-of-sample fit, so we use it in our main specifications. But, the LASSO increases out-of-sample fit by roughly $\Delta\bar{R}_n^2 = 1.2\%$ pt regardless of the number of lags used in the benchmark model. In fact, the row labeled “AR(h^*)” shows that the LASSO still increases out-of-sample fit by $\Delta\bar{R}_n^2 = 1.1\%$ pt relative to an autoregressive benchmark model where the number of lags has been optimally chosen within each 30-minute estimation window using the Bayesian information criteria.

Including Portfolio Returns. Finally, the last three rows in Table 2 show that the LASSO’s increase in out-of-sample fit persists even when the benchmark model includes lagged industry and risk-factor returns. In particular, the last row examines the LASSO’s increase in out-of-sample fit relative to a benchmark model that includes 3 lags of the market’s returns, 3 lags of a stock’s industry’s returns, 3 lags of a size portfolio’s returns, 3 lags of a value portfolio’s returns, and 3 lags of a stock’s

own lagged returns. Using this benchmark model to generate 1-minute-ahead return forecasts means estimating an OLS regression with $1 + 5 \times 3 = 16$ coefficients in just 30 minutes of data. The fact that this more elaborate benchmark has a lower out-of-sample fit than the simple AR(3) benchmark is evidence of overfitting. By contrast, the fact that using the LASSO increases out-of-sample fit relative to this same AR(3) benchmark suggests the LASSO is capturing meaningful new information.

Cross-Sectional Variation. The LASSO’s increase in out-of-sample fit does not appear to be explained by observable differences across stocks. For example, Figure 3 shows that the LASSO’s $\Delta \bar{R}_n^2 = 1.185\%$ pt increase relative to the AR(3) benchmark and its $\Delta \bar{R}_n^2 = 2.469\%$ pt increase relative to the market benchmark are remarkably consistent across industries. In addition, Table 3 shows that the LASSO’s increase in out-of-sample fit is not explained by firm characteristics such as market capitalization, trading volume, return volatility, or bid-ask spread. In fact, the LASSO actually increases out-of-sample fit slightly more for large, liquid, frequently traded stocks—that is, stocks with returns that are the hardest to predict using conventional methods.

Time-Series Variation. How does the LASSO’s increase in out-of-sample fit vary over time? The left panel of Figure 4 shows that the LASSO increased in out-of-sample fit less during the financial crisis. This pattern is consistent with existing research showing that returns display a stronger single-factor structure during market downturns (Ferson and Harvey, 1991; Dangl and Halling, 2012; Kacperczyk et al., 2014). If there is a strong single-factor structure, then there cannot be as many unexpected, short-lived, and sparse signals in the cross-section of returns. And, if there are not many unexpected, short-lived, and sparse signals in the cross-section of returns, then using the LASSO should not increase out-of-sample fit by much.

The right panel of Figure 4 looks at the intraday time-series properties of the LASSO’s increase in out-of-sample fit. To create this figure, we re-estimated the regressions in Equations (12) and (13) 356 times each trading day using the cross-section of all 250 forecasts. We then averaged the resulting adjusted R^2 statistics in each minute across trading days. The fact that the LASSO’s increase in out-of-sample fit is so smooth suggests that our findings are not explained by intraday microstructure effects.

Shrinkage Without Selection. As noted earlier, the term “LASSO” is an acronym standing for the Least Absolute Shrinkage and Selection Operator. And, as suggested by this name, the LASSO’s penalty function is performing two separate tasks. First,

Increase in Out-of-Sample Fit, Main Results

	$\bar{R}_n^{2,\text{Bmk}}$ [%]	$\Delta\bar{R}_n^2$ [%pt]	p -val.
AR(3)	7.365 (0.076)	1.185 [1.162, 1.208] (0.012)	0.000
Market	0.311 (0.003)	2.469 [2.416, 2.522] (0.027)	0.000
AR(3), Market	5.553 (0.058)	1.424 [1.395, 1.453] (0.015)	0.000
AR(1)	6.061 (0.052)	1.288 [1.263, 1.314] (0.013)	0.000
AR(2)	7.309 (0.071)	1.165 [1.143, 1.188] (0.012)	0.000
AR(4)	7.174 (0.076)	1.238 [1.214, 1.262] (0.012)	0.000
AR(5)	6.725 (0.074)	1.307 [1.282, 1.332] (0.013)	0.000
AR(h^*)	8.031 (0.080)	1.134 [1.113, 1.156] (0.011)	0.000
Market, Industry	0.314 (0.007)	2.436 [2.384, 2.489] (0.027)	0.000
Market, Size, Value	0.214 (0.003)	2.443 [2.390, 2.496] (0.027)	0.000
AR(3), Market, Industry, Size, Value	1.443 (0.015)	2.232 [2.184, 2.279] (0.024)	0.000

Table 2: *The LASSO’s increase in out-of-sample fit relative to a variety of benchmark models measured as the percentage point increase in adjusted R^2 . (Sample) Regression results for a randomly selected subset of 250 stocks on each trading day from January 2005 to December 2012. (Estimates) $\bar{R}_n^{2,\text{Bmk}}$: Out-of-sample fit of a benchmark model measured using the adjusted R^2 statistic. $\Delta\bar{R}_n^2$: The LASSO’s increase in out-of-sample fit. p -val.: Probability of observing the realized increase in out-of-sample fit $\Delta\bar{R}_n^2$ under null hypothesis of no increase. Numbers in parentheses are standard errors clustered by stock-day. Numbers in square brackets represent the 95% confidence interval for the mean. (Benchmark Models) AR(3): 3 lags of a stock’s own returns. Market: 3 lags of the market’s returns. AR(3), Market: 3 lags of a stock’s own returns and 3 lags of the market’s returns. AR(1): 1 lag of a stock’s own returns. AR(2): 2 lags of a stock’s own returns. AR(4): 4 lags of a stock’s own returns. AR(5): 5 lags of a stock’s own returns. AR(h^*): h^* lags of a stock’s own returns where h^* is chosen within each 30-minute estimation window using the Bayesian information criteria. Market, Industry: 3 lags of the market’s returns and 3 lags of the stock’s industry’s returns. Market, Size, Value: 3 lags of the market’s returns, a size portfolio’s returns, and a value portfolio’s returns respectively. AR(3), Market, Industry, Size, Value: 3 lags of the market’s returns, a stock’s industry’s returns, a size portfolio’s returns, and a value portfolio’s returns respectively. (Reads) “A researcher could explain an additional $\Delta\bar{R}_n^2 = 1.185\%$ pt of the variation in returns by using both the LASSO and an AR(3) model rather than just the AR(3) model alone.”*

it is selecting coefficients larger than λ_n . In the top panel of Table 4, we study the role of selection in the LASSO’s increase in out-of-sample fit. We do this by examining the LASSO’s increase in out-of-sample fit relative to alternative benchmark models that already incorporate some form of shrinkage. If the LASSO’s increase in out-of-sample fit were only due to coefficient shrinkage and not due to selection, then using the LASSO in addition to one of these benchmark models should yield $\Delta\bar{R}_n^2 = 0\%$ pt.

We consider two different kinds of pure-shrinkage benchmarks. To start with, we compute 1-minute-ahead return forecasts using Ridge (Hastie et al., 2001, §3.4). The first row shows that shrinkage on its own actually results in a lower out-of-sample fit. And, the LASSO increases out-of-sample fit relative to this pure-shrinkage benchmark.

The second row contains a weighted average of the 1-minute-ahead return forecasts generated by the AR(3) and market benchmarks as suggested in Rapach et al. (2010). Consistent with their original results, we find that averaging these forecasts generates a higher out-of-sample fit than including each of the forecasts separately, $\bar{R}_n^{2,\text{Bmk}} = 6.210\%$ vs. $\bar{R}_n^{2,\text{Bmk}} = 5.553\%$. But, the LASSO still increases out-of-sample fit relative to the weighted-average benchmark by $\Delta\bar{R}_n^2 = 1.302\%$ pt.

Selection Without Shrinkage. Second, the LASSO is shrinking the point estimates of all selected coefficients by λ_n . The middle panel investigates whether this shrinkage component is key to the LASSO’s increase in out-of-sample fit. We do this by examining whether LASSO alternatives, which do not use any sort of shrinkage, increase out-of-sample fit by more than the LASSO does.

Again, we take two different approaches. The first row shows that using an OLS regression that only includes the single predictor with the highest in-sample correlation performs much worse than the LASSO. This is tantamount to applying a forward-stepwise procedure to the standard OLS-regression framework. And, the second row shows that using a two-step procedure that fits an OLS regression to the predictors selected by the LASSO generates the same increase in out-of-sample fit as just using the LASSO itself. This suggests that there are not substantial gains to be made by post-selection inference. It also implies that using an alternative sparse-selection procedure, such as the adaptive LASSO (Zou, 2006) or elastic net (Zou and Hastie, 2005), will not qualitatively affect our results because these procedures only differ from the LASSO in the amount of shrinkage.

Lambda Noise. Finally, recall that our implementation of the LASSO selects the

Increase in Out-of-Sample Fit, by Industry

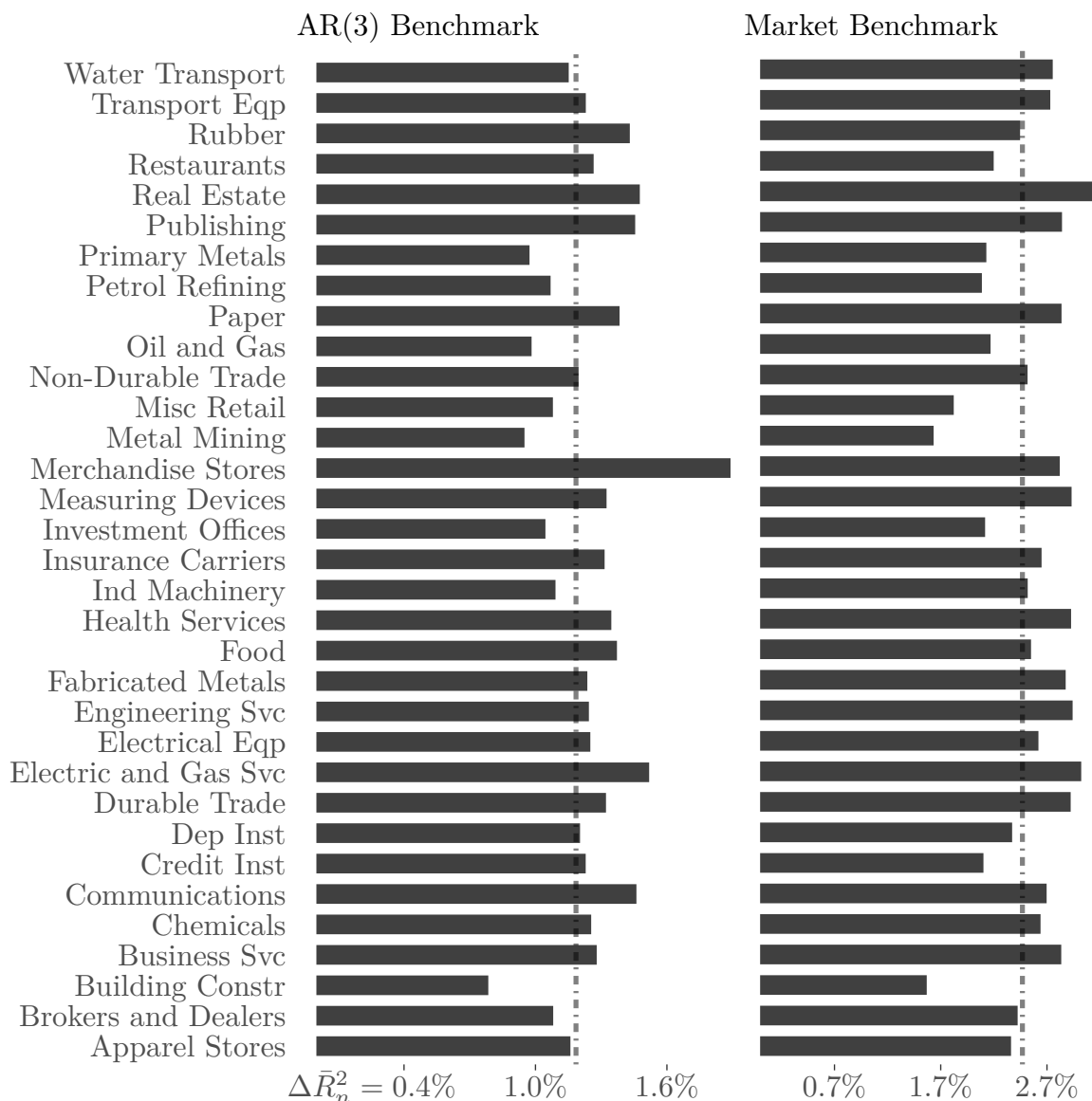


Figure 3: *The LASSO’s increase in out-of-sample fit relative to the AR(3) and market benchmarks sorted by 3-digit SIC industries. Increase in out-of-sample fit is measured as the percentage point increase in adjusted R^2 . (Sample) Regression results for a randomly selected subset of 250 stocks on each trading day from January 2005 to December 2012. We restrict the sample to industries with at least 20 firms on average. (Left) The LASSO’s increase in out-of-sample fit relative to a AR(3) benchmark with the dashed line representing $\Delta\bar{R}_n^2 = 1.185\%pt.$ (Right) The LASSO’s increase in out-of-sample fit relative to a market benchmark with the dashed line representing $\Delta\bar{R}_n^2 = 2.469\%pt.$ (Reads) “The LASSO increases out-of-sample fit for all industries.”*

Increase in Out-of-Sample Fit, by Characteristics

AR(3) Benchmark		High, > 50%ile		Low, < 50%ile		
		$\Delta \bar{R}_n^2$ [%pt]	<i>p</i> -val.	$\Delta \bar{R}_n^2$ [%pt]	<i>p</i> -val.	<i>p</i> -val.
Mkt Cap	1.341 (0.017)	[1.307, 1.375]	0.000	1.029 (0.011)	[1.007, 1.051]	0.000
Volume	1.347 (0.017)	[1.315, 1.380]	0.000	1.023 (0.011)	[1.002, 1.044]	0.000
Volatility	1.137 (0.012)	[1.113, 1.162]	0.000	1.234 (0.016)	[1.204, 1.265]	0.000
Spread	0.996 (0.011)	[0.975, 1.017]	0.000	1.377 (0.017)	[1.343, 1.410]	0.000

Market Benchmark		High, > 50%ile		Low, < 50%ile		
		$\Delta \bar{R}_n^2$ [%pt]	<i>p</i> -val.	$\Delta \bar{R}_n^2$ [%pt]	<i>p</i> -val.	<i>p</i> -val.
Mkt Cap	2.605 (0.036)	[2.534, 2.675]	0.000	2.332 (0.027)	[2.280, 2.385]	0.000
Volume	2.518 (0.033)	[2.454, 2.582]	0.000	2.419 (0.030)	[2.360, 2.478]	0.000
Volatility	2.680 (0.028)	[2.626, 2.734]	0.000	2.259 (0.032)	[2.196, 2.321]	0.000
Spread	2.302 (0.026)	[2.251, 2.353]	0.000	2.639 (0.036)	[2.569, 2.709]	0.000

Table 3: *The LASSO's increase in out-of-sample fit relative to the AR(3) and market benchmarks sorted by firm characteristics. Increase in out-of-sample fit is measured as the percentage point increase in adjusted R^2 . (Sample) Regression results for a randomly selected subset of 250 stocks on each trading day from January 2005 to December 2012. (Estimates) $\Delta \bar{R}_n^2$: The LASSO's increase in out-of-sample fit relative a benchmark model. *p*-val.: Probability of observing the realized $\Delta \bar{R}_n^2$ under the null hypothesis of no increase. Numbers in parentheses are standard errors clustered by stock-day. Numbers in square brackets are 95% confidence intervals. (Characteristics) Mkt Cap: Market value at close of previous trading day. Volume: Trading volume on previous trading day. Volatility: Volatility of 1-minute returns during previous trading day. Spread: Average bid-ask spread during previous trading day. High, > 50%ile: Subset of stocks with above-median values for a given characteristic. Low, < 50%ile: Subset of stocks with below-median values for the same characteristic. (Reads) "The LASSO increases out-of-sample fit slightly more for large, liquid, frequently-traded stocks."*

Increase in Out-of-Sample Fit, Source of Gains

No Selection	$\bar{R}_n^{2,\text{Bmk}}$ [%]	$\Delta\bar{R}_n^2$ [%pt]	p-val.
Ridge	1.182 (0.012)	2.447 [2.395, 2.498] (0.026)	0.000
Average OLS	6.210 (0.065)	1.302 [1.276, 1.328] (0.013)	0.000

No Shrinkage	AR(3) Benchmark			Market Benchmark		
	$\Delta\bar{R}_n^2$ [%pt]		p-val.	$\Delta\bar{R}_n^2$ [%pt]		p-val.
Single-Best Predictor	0.474 [0.464, 0.485] (0.005)		0.000	0.671 [0.658, 0.685] (0.007)		0.000
Post-Selection OLS	1.245 [1.218, 1.271] (0.013)		0.000	2.544 [2.483, 2.604] (0.031)		0.000

Lambda Noise	AR(3) Benchmark			Market Benchmark		
	$\Delta\bar{R}_n^2$ [%pt]		p-val.	$\Delta\bar{R}_n^2$ [%pt]		p-val.
LASSO, $\lambda_n^{+s.e}$	0.829 [0.807, 0.851] (0.011)		0.000	1.555 [1.518, 1.592] (0.019)		0.000

Table 4: Source of the LASSO's increase in out-of-sample fit relative to the AR(3) and market benchmarks. Increase in out-of-sample fit is measured as the percentage point increase in adjusted R^2 . **(Sample)** Regression results for a randomly selected subset of 250 stocks on each trading day from January 2005 to December 2012. **(Columns)** $\bar{R}_n^{2,\text{Bmk}}$: Out-of-sample fit of benchmark model. $\Delta\bar{R}_n^2$: Increase in out-of-sample fit for the LASSO or alternative procedure. p-val.: Probability of observing the realized $\Delta\bar{R}_n^2$ under null hypothesis of no increase in out-of-sample fit. Numbers in parentheses are standard errors clustered by stock-day. Numbers in square brackets are 95% confidence intervals. **(No Selection)** The LASSO's increase in out-of-sample fit relative to alternative benchmark models that involve shrinkage but no selection. Ridge: Ridge regression estimated using 3 lags. Average OLS: Weighted average of AR(3) and market benchmark models. **(No Shrinkage)** Increase in out-of-sample fit for alternative procedures that involve selection but no shrinkage. Single-Best Predictor: OLS regression using the single predictor with the highest correlation during the 30-minute estimation window. Post-Selection OLS: OLS regression using predictors selected by the LASSO. **(Lambda Noise)** The LASSO's increase in out-of-sample fit when using a penalty parameter that is 1-standard-deviation too large.

Increase in Out-of-Sample Fit, Time Series

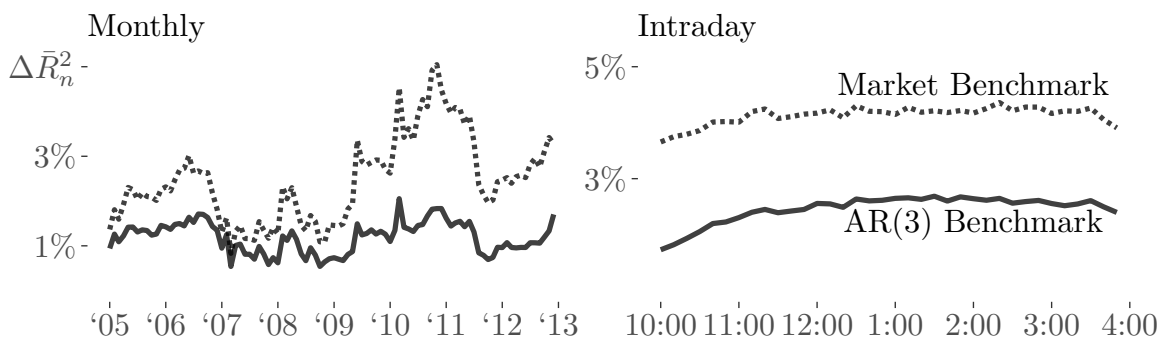


Figure 4: *Time-series variation in the LASSO’s increase in out-of-sample fit relative to the AR(3) and market benchmarks. Increase in out-of-sample fit is measured as the percentage point increase in adjusted R^2 . (Sample) Regression results for a randomly selected subset of 250 stocks on each trading day from January 2005 to December 2012. (Monthly) The LASSO’s increase in out-of-sample fit by month. (Intraday) The LASSO’s increase in out-of-sample fit by minute of the trading day. (Reads) “The LASSO increased in out-of-sample fit less during the financial crisis, but the LASSO’s increase in out-of-sample fit is relatively stable over the course of the trading day.”*

penalty parameter, λ_n , for each stock within each 30-minute estimation window using K -fold cross validation. This means that our choice of λ_n contains some estimation error. The bottom panel of Table 4 examines the effect of this estimation error on our results. We do this by testing whether the LASSO still increases out-of-sample fit even when we use a penalty parameter that is 1-standard-deviation too large in terms of the optimal λ_n ’s cross-validation error. We find that, while using $\lambda_n^{+s.e.}$ rather than the optimal λ_n reduces the LASSO’s gains, it does not eliminate them.³

3.4 Sharpe Ratios

We have just seen that using the LASSO leads to a statistically significant 1.2%pt increase in out-of-sample fit. But, what is the economic magnitude of this gain? Because “any predictive regression can be expressed as a portfolio sort (Pedersen,

³We only look at alternative values of λ_n that are larger than the optimal λ_n . Here is why. If you take $\lambda_n \rightarrow \infty$, then the LASSO ignores all predictors and simply estimates the mean. So, conditional on the LASSO converging at the optimal λ_n , any alternative choice for λ_n that’s larger than this optimal value will deliver well-defined parameter estimates. However, shrinking $\lambda \rightarrow 0$ means including more predictors. So, alternative values of λ_n that are smaller than the best-fit penalty parameter might allow for more than 30 predictors, which would yield parameter estimates that are not well defined (i.e., which will depend on the details of the optimization algorithm used).

2015, p. 51),” we answer this question by studying the performance of a LASSO-implied trading strategy.

Naïve Starting Point. Here is the starting point for our LASSO-implied strategy. Note that the slope coefficient, \bar{b}_n , from Equation (11) can be written as:

$$\bar{b}_n \stackrel{\text{def}}{=} \frac{\overline{\text{Cov}}\left[r_{n,t+1}, \frac{f_{n,t}^{\text{LASSO}} - \bar{m}_n^{\text{LASSO}}}{\bar{s}_n^{\text{LASSO}}}\right]}{\overline{\text{Var}}\left[\frac{f_{n,t}^{\text{LASSO}} - \bar{m}_n^{\text{LASSO}}}{\bar{s}_n^{\text{LASSO}}}\right]} \quad (14)$$

In the equation above, $r_{n,t+1}$ is the n th stock’s realized return in minute $(t + 1)$, $f_{n,t}^{\text{LASSO}}$ is the LASSO’s 1-minute-ahead return forecast for minute $(t + 1)$, \bar{m}_n^{LASSO} and \bar{s}_n^{LASSO} denote the mean and standard deviation of the LASSO’s 1-minute-ahead return forecasts for the n th stock over the course of the entire trading day, and the operators $\overline{\text{Cov}}[\cdot]$ and $\overline{\text{Var}}[\cdot]$ denote the covariance and variance over the same time period—that is, one observation for each minute from 10:04am to 3:59pm.

Because we normalize the LASSO’s 1-minute-ahead return forecasts to have mean zero and unit variance within a given trading day, the denominator in Equation (14) is equal to one. Thus, we can write out the expression for \bar{b}_n using just the covariance term as follows

$$\bar{b}_n = \overline{\text{Cov}}\left[r_{n,t+1}, \frac{f_{n,t}^{\text{LASSO}} - \bar{m}_n^{\text{LASSO}}}{\bar{s}_n^{\text{LASSO}}}\right] \quad (15a)$$

$$= \frac{1}{356} \cdot \sum_{h=1}^{356} (r_{n,t+h} - \bar{a}_n) \cdot \left(\frac{f_{n,t+h-1}^{\text{LASSO}} - \bar{m}_n^{\text{LASSO}}}{\bar{s}_n^{\text{LASSO}}}\right) \quad (15b)$$

where \bar{a}_n denotes the n th stock’s mean return during the day.

This expression can be interpreted as an average return to a naïve LASSO-implied strategy:

$$r_{n,t+1}^{\text{Naïve}} = (r_{n,t+1} - \bar{a}_n) \times (1/\bar{s}_n^{\text{LASSO}}) \times (f_{n,t}^{\text{LASSO}} - \bar{m}_n^{\text{LASSO}}) \quad (16)$$

The strategy is long $1/\bar{s}_n^{\text{LASSO}}$ shares of the n th stock for each 1%pt that the LASSO’s 1-minute-ahead return forecast exceeds its mean. Likewise, it is short $1/\bar{s}_n^{\text{LASSO}}$ shares of the n th stock for each 1%pt that the LASSO’s forecast falls below its mean for the day. The strategy will be profitable if the n th stock’s realized return in minute $(t + 1)$ tends to be above average, $r_{n,t+1} > \bar{a}_n$, when the LASSO’s return forecast is also above average, $f_{n,t}^{\text{LASSO}} > \bar{m}_n^{\text{LASSO}}$, and vice versa. Thus, we can interpret Table 1 as saying that, for a randomly selected stock-day, this naïve LASSO-implied strategy has a gross return of $\bar{b}_n = 1.433\%$ per month.

There are three problems with implementing this naïve LASSO-implied strategy, though. To start with, the strategy ignores trading costs (Hasbrouck, 2009; Frazzini et al., 2015; Novy-Marx and Velikov, 2015), which are substantial when trading at the 1-minute horizon. The strategy also suffers from look-ahead bias. We do not know the n th stock’s mean return, \bar{a}_n , or the mean and variance of the LASSO’s 1-minute-ahead return forecasts, \bar{m}_n^{LASSO} and \bar{s}_n^{LASSO} , until the end of the trading day. Finally, the strategy gives no guidance on how we should combine the 250 different 1-minute-ahead return forecasts we make each minute. So, we modify the strategy in Equation (16) in several ways to account for these problems.

Trading Costs. Let’s start with the trading-cost adjustments. To account for trading costs, we make two changes. First, we change the naïve LASSO-implied strategy so that it only executes an order when the LASSO’s 1-minute-ahead return forecast exceeds the bid-ask spread, $1_{\{|f_{n,t}^{\text{LASSO}}| > \text{sprd}_{n,t}\}}$. Then, conditional on a trade taking place, we compute the strategy’s realized returns net of this bid-ask spread, $(r_{n,t+1} - \bar{a}_n) \cdot \text{Sign}[f_{n,t}^{\text{LASSO}}] - \text{sprd}_{n,t}$.

Look-Ahead Bias. Next, let’s turn to the issue of look-ahead bias. To eliminate any look-ahead bias, our strategy should only use values for \bar{a}_n , \bar{m}_n^{LASSO} , and \bar{s}_n^{LASSO} that are known prior to trading in minute $(t + 1)$. To make sure this is the case, we start by assuming that both realized returns and 1-minute-ahead return forecasts are mean-zero variables, $\bar{a}_n = \bar{m}_n^{\text{LASSO}} = 0$. This is a reasonable assumption at the 1-minute horizon as emphasized by our point estimate of $\bar{a}_n = 0.002\%$ per month in Table 1. In addition, we also estimate the standard deviation of the LASSO’s 1-minute-ahead return forecast during the previous 30 minutes of trading rather than over the course of the entire trading day. We use $\tilde{\sigma}_{n,t}^{\text{LASSO}}$ to denote this in-sample standard-deviation estimate.

Portfolio Construction. Finally, we come to the question of portfolio construction: how do we combine the 250 different forecasts each minute into a single portfolio? The naïve strategy buys/sells more shares of the n th stock per 1%pt change in the LASSO’s forecast when the LASSO’s 1-minute-ahead return forecasts for the n th stock tend to be less volatile throughout the trading day. Intuitively, a 1-minute-ahead return forecast that is 0.20%pt above/below average is a very different signal when the typical return forecast is on the order of $\bar{m}_n^{\text{LASSO}} \pm 0.02\%$ pt than when it is on the order of $\bar{m}_n^{\text{LASSO}} \pm 2.00\%$ pt. In the first case, a 0.20%pt above/below average forecast is noteworthy, garnering a large portfolio weight; whereas, in the second case,

it is commonplace, garnering a small portfolio weight.

We want our strategy to have a similar flavor. But, we also want to be able to fund the strategy by borrowing \$1 at the riskless rate prior to the beginning of each trading day. Then, the strategy's returns at the end of the trading day minus the riskless rate would represent the excess returns to a zero-cost portfolio. The complication comes from the fact that the number of LASSO forecasts that exceed the spread, $1_{\{|f_{n,t}^{\text{LASSO}}| > \text{sprd}_{n,t}\}}$, can change from minute to minute. So, we need a rule that determines how much we invest for each of the stocks with a forecast that exceeds the spread in a given minute.

Here is how we do this. Suppose that the LASSO generated a 1-minute-ahead return forecast for the n th stock that exceeds the spread in minute t . Then, the LASSO-implied strategy invests a fraction $\tilde{\omega}_{n,t}^{\text{LASSO}}$ of its assets in the n th stock:

$$\tilde{\omega}_{n,t}^{\text{LASSO}} \stackrel{\text{def}}{=} \frac{(1/\tilde{\sigma}_{n,t}^{\text{LASSO}}) \cdot f_{n,t}^{\text{LASSO}} \cdot 1_{\{|f_{n,t}^{\text{LASSO}}| > \text{sprd}_{n,t}\}}}{\sum_{n'=1}^{250} \left| (1/\tilde{\sigma}_{n',t}^{\text{LASSO}}) \cdot f_{n',t}^{\text{LASSO}} \cdot 1_{\{|f_{n',t}^{\text{LASSO}}| > \text{sprd}_{n',t}\}} \right|} \quad (17)$$

The denominator is the total investment, both long and short, dictated by the naïve strategy in stocks with forecasts that exceeded the spread in minute t . The numerator is the investment dictated by the naïve strategy in the n th stock. The settlement period for NYSE stocks is longer than 1 minute, so it does not make sense to net-out offsetting short and long positions at the 1-minute horizon. This is why we use the absolute-value operator in the denominator. If there are no stocks with LASSO forecasts that exceed the spread in a given minute, then $\tilde{\omega}_{n,t}^{\text{LASSO}} = 0$ for all stocks.

Consider a quick example to see what these weights imply. Suppose that there are 8 stocks with LASSO forecasts that exceed the spread. And, to keep things simple, suppose that the first 4 forecasts are +1%pt while the second 4 forecasts are -1%pt. If all 8 forecasts have a volatility of 0.02%, then the denominator will be $8 \cdot |(1/0.02\%) \cdot 1\%pt| = 400$. And, the LASSO-implied strategy will invest 1/8th of its portfolio position in each stock. But, if stocks 1, 2, 5, and 6 instead have forecast volatilities of 2%, then the denominator will be $4 \cdot |(1/0.02\%) \cdot 1\%pt| + 4 \cdot |(1/2\%) \cdot 1\%pt| = 202$. This means that stocks 1, 2, 5, and 6 will have portfolio weights of $\tilde{w}_{n,t}^{\text{LASSO}} = \pm 1/404$ while stocks 3, 4, 7, and 8 will have portfolio weights of $\tilde{w}_{n,t}^{\text{LASSO}} = \pm 100/404$ —that is, weights that are 100-times larger in magnitude.

LASSO-Implied Strategy. Thus, the net return each minute to our LASSO-implied

strategy is:

$$r_{t+1}^{\text{LASSO}} = \sum_{n=1}^{250} \{r_{n,t+1} \cdot \text{Sign}[f_{n,t}^{\text{LASSO}}] - \text{sprd}_{n,t}\} \times \tilde{\omega}_{n,t}^{\text{LASSO}} \quad (18)$$

This strategy accounts for trading costs and does not suffer from look-ahead bias. In addition, if we subtract off the riskless rate from this strategy’s end-of-day returns, then the resulting excess returns correspond to the returns of a zero-cost portfolio funded by borrowing \$1 at the beginning of the trading day.

Sharpe Ratios. Table 5 compares the performance of this LASSO-implied strategy with other benchmarks. Let’s start with the top panel, which provides information on annualized Sharpe ratios. The first column reports the Sharpe ratio of an unconditional buy-and-hold S&P 500 strategy. This strategy generates an annualized Sharpe ratio of only 0.123. This Sharpe ratio is relatively low because the S&P 500’s returns were quite volatile during our sample period, which includes the financial crisis. The next column reports the LASSO-implied strategy’s Sharpe ratio, which was 1.791. This point estimate is more than 12-times larger than the Sharpe ratio of the unconditional buy-and-hold strategy.

The last column describes the performance of an AR(3)-implied strategy that is analogous to the LASSO-implied strategy—that is, a strategy constructed by replacing $f_{n,t}^{\text{LASSO}}$ with $f_{n,t}^{\text{AR}(3)}$ in Equation (18). It turns out that this AR(3)-implied strategy loses money. Even though the AR(3) benchmark had higher out-of-sample fit than the LASSO, it is not possible to trade on the AR(3) benchmark’s 1-minute-ahead return forecasts after adjusting for the bid-ask spread and look-ahead bias.

Thus, the $\bar{R}_n^{2,\text{Bmk}} = 7.365\%$ out-of-sample fit for the AR(3) benchmark does not translate into economically meaningful performance. This finding is broadly consistent with the results in both [Korajczyk and Sadka \(2004\)](#) and [Patton and Weller \(2017\)](#), who give evidence that momentum profits disappear after adjusting for trading costs. The forecast-implied strategy associated with the market benchmark performs even worse than the one associated with the AR(3) benchmark.

Abnormal Returns. The positive excess returns earned by the LASSO-implied strategy do not appear to be explained by standard risk factors. In the bottom panel of Table 5, we study the LASSO-implied strategy’s covariance with the daily return on the market, the daily returns on size and value portfolios ([Fama and French, 1993](#)), and the daily return on a momentum portfolio ([Carhart, 1997](#)). Because the LASSO-

implied strategy is zero cost, we choose the scale of the initial investment so that it has the same average excess return as the buy-and-hold S&P 500 strategy, 2.719% per year. Because the net returns to the LASSO-implied strategy are so smooth—12 times less volatile than the returns on the S&P 500—we find that exposure to these commonly-used risk factors at the daily horizon explains almost none of the LASSO-implied strategy’s excess returns.

Increases vs. Levels. If the AR(3)-implied strategy had delivered positive excess returns, then we would have examined whether a 50/50 combination of the LASSO- and AR(3)-implied strategy produced an even higher Sharpe ratio than the AR(3)-implied strategy on its own (Asness et al., 2013; Novy-Marx, 2016). But, such a test is no longer necessary given that the AR(3)-implied strategy loses money. If the AR(3)-implied strategy delivers negative net returns, then there is no way that it could be explaining the LASSO-implied strategy’s positive net returns. The difference in signs implies a difference in the signals being used.

Trade Frequency. Finally, Table 6 gives summary statistics describing the rate at which both the LASSO- and AR(3)-implied strategies trade. The first row of the top panel reveals that the AR(3)-implied strategy trades roughly twice as often as the LASSO-implied strategy, 17.643 trades vs. 8.624 trades per minute out of 250 possible trades. The second row shows that roughly half of each strategy’s trades are buy orders, and the third row shows that roughly a third of each strategy’s trades are profitable. The bottom panel then shows that the LASSO-implied strategy is more active in large, liquid, frequently traded stocks.

4 Predictor Analysis

After seeing the increase in out-of-sample fit associated with the LASSO, a natural next question is: What do the predictors selected by the LASSO look like? In this section, we give evidence that the LASSO identifies predictors that are unexpected, short-lived, and sparse.

4.1 Unexpected

There are many lead-lag relationships already documented in the academic literature at the weekly and monthly horizons. For example, Lo and MacKinlay (1990) and

Forecast-Implied Performance Net of Trading Costs

Annualized Sharpe Ratios		
S&P 500	LASSO	AR(3)
0.123	1.791	−0.662

LASSO-Implied Strategy Abnormal Returns [%/yr]	α	Mkt	HmL	SmB	Mom
Market	2.709 (0.034)	0.004 (0.002)			
3-Factor Model	2.713 (0.034)	0.004 (0.002)	−0.004 (0.004)	0.000 (0.003)	
4-Factor Model	2.707 (0.034)	0.005 (0.002)	−0.004 (0.004)	0.003 (0.004)	0.003 (0.004)

Table 5: *Performance of forecast-implied strategies net of trading costs. (Sample) Each trading day from January 2005 to December 2012. (Sharpe Ratios) Annualized Sharpe ratios of forecast-implied trading strategies net of trading costs. First column reports results for strategy that invests \$1 in the S&P 500 at market open on January 3rd, 2005 and holds that position until market close on December 31st, 2012. The next column reports results for the LASSO’s forecast-implied trading strategy over the same time period. The third column reports analogous results for an AR(3)-implied strategy with the same initial investment. (Abnormal Returns) Net abnormal returns of the LASSO-implied strategy relative to the market, the Fama and French (1993) 3-factor model, and the Carhart (1997) 4-factor model. The size of the initial investment in the LASSO-implied strategy was chosen so that it has the same average excess return as the buy-and-hold S&P 500 strategy. First column reports annualized abnormal returns. Remaining columns report dimensionless slope coefficients associated with each factors. (Reads) “The LASSO-implied strategy generates positive excess returns net of the spread with an annualized Sharpe ratio of 1.791, and these excess returns are not explained by the strategy’s exposures to standard risk factors.”*

Trading Frequency of Forecast-Implied Strategies

Aggregate [# /min]	LASSO	AR(3)
Trades	8.624	17.643
Buy Orders	4.306	8.887
Successful	2.600	6.538

By Characteristics [# /min]	>50%ile Mkt Cap	>50%ile Volume	>50%ile Volatility	<50%ile Spread
LASSO Trades	5.188	5.115	4.821	5.114
LASSO Buy Orders	2.592	2.555	2.407	2.556
LASSO Successful	1.578	1.533	1.467	1.550

Table 6: Trading frequency of forecast-implied strategies. **(Sample)** Minute-level trades for randomly selected subset of 250 stocks on each trading day from January 2005 to December 2012 for which we compute 1-minute-ahead return forecasts using the LASSO. **(Aggregate)** First row reports the average number of trades per minute made by LASSO- and AR(3)-implied strategies. Second row reports the number of buy orders per minute. And, third row reports the number of trades per minute that made money after accounting for the bid-ask spread. **(By Characteristics)** Number of trades, buy orders, and successful trades per minute for the LASSO-implied strategy among large stocks, stocks with high trading volume, stocks with high return volatility, and liquid stocks. **(Reads)** “The AR(3)-implied strategy trades roughly twice as often as the LASSO-implied strategy. And, the LASSO-implied strategy is more active in large, liquid, frequently traded stocks.”

DeMiguel et al. (2014) both show that the returns of large stocks predict the future returns of small stocks, Chordia and Swaminathan (2000) shows that the returns of high-volume stocks predict the future returns of low-volume stocks, and Hou (2007) shows that the returns of larger stocks within an industry predict the future returns of smaller stocks within the same industry. We now give evidence that the LASSO’s choice of predictors does not follow these existing lead-lag relationships—in other words, evidence that the LASSO identifies an unexpected collection of predictors.

Econometric Approach. Suppose that stock n is one of the randomly chosen 250 stocks for which we use the LASSO to make 1-minute-ahead return forecasts on day d . For each stock $n' \in \{1, \dots, N\}$, we define an indicator variable that is one if stock

n' was ever used by the LASSO as a predictor in any of the 356 forecasts for stock n on day d and zero otherwise. Note that for computational reasons this indicator variable summarizes the contribution of all 3 lags of the n' th stock's returns. We then use a logit regression with this indicator as the dependent variable to study how the characteristics of stock n' are related to the probability that the LASSO identifies its lagged returns as a predictor for stock n at some point during the trading day.

Empirical Results. Table 7 contains the results of these logit regressions. The first column shows that the LASSO is no more likely to select large stocks than small stocks as predictors. The second column shows that infrequently traded stocks are just as likely to be selected as predictors by the LASSO as frequently traded stocks. The third column shows that return volatility is not related to the LASSO's choice of predictors. And, the fourth column shows that the LASSO is equally likely to use liquid and illiquid stocks as predictors. In short, these four columns document a lack of correlation between firm characteristics and the LASSO's choice of predictors. The fifth column, by contrast, reveals that the LASSO is 1.15% *less* likely to select a stock from the same industry as a predictor. Thus, the LASSO is not just selecting predictors based off of the intra-industry lead-lag effect documented in Hou (2007). The final column shows that all of these patterns hold up when these variables are included in the same regression specification.

Penalty Parameter. How can the LASSO increase out-of-sample fit at the 1-minute horizon but still be so unrelated to firm characteristics? After all, researchers know that these characteristics are correlated with returns at the weekly and monthly horizons. This seeming contradiction resolves itself when we examine the size of the LASSO's penalty parameter, λ_n . The LASSO sets all coefficients smaller than λ_n to be precisely zero—that is, it completely ignores these predictors.

Figure 5 plots the mean and 95% confidence interval for λ_n each month during our sample period. The horizontal dashed line denotes the sample average for λ_n , which is 2.5% per month. In other words, the LASSO typically ignores all predictors weaker than $\lambda = 2.5\%$ per month when making its 1-minute-ahead return forecasts. This lower bound is more than twice as large as the size of well-known predictors at the weekly and monthly horizon (McLean and Pontiff, 2016; Harvey et al., 2016; Linnainmaa and Roberts, 2016).

This is why our implementation of the LASSO at the 1-minute horizon is not identifying the existing steady long-lived predictors documented in the academic lit-

Characteristics of LASSO Predictors

	Dep. Variable: Ever Selected	
Mkt Cap > 50%ile	0.005 (0.004) [0.11%]	0.003 (0.007) [0.07%]
Volume > 50%ile	0.004 (0.004) [0.09%]	0.001 (0.006) [0.01%]
Volatility > 50%ile	−0.005 (0.004) [0.13%]	−0.004 (0.004) [0.10%]
Spread < 50%ile	0.002 (0.004) [0.05%]	0.006 (0.006) [0.15%]
In Same Industry	−0.046 (0.006) [1.15%]	−0.046 (0.006) [1.14%]

Table 7: *Characteristics of the predictors selected by the LASSO. (Sample) On each trading day, d , from January 2005 to December 2012 the data contain one observation per predictor for each of the randomly selected 250 stocks for which we make 1-minute-ahead return forecasts. e.g., on October 6th, 2010 the LASSO could choose as predictors any of the $N = 2,191$ NYSE-listed stocks. So, on that day, our data contained $250 \times 2,191 = 547,750$ separate observations. (Specification) Each column reports the results from a different logit regression. Point estimates are log-odds ratios. Numbers in parentheses are standard errors clustered by day and predictor (i.e., stock n'). Numbers in square brackets are the marginal effects implied by the log-odds ratios. (Variables) Dep. Variable: An indicator variable that is one if stock n' was ever used by the LASSO as a predictor when making 1-minute-ahead return forecasts for stock n on day d . Mkt Cap > 50%ile: An indicator variable that is one if stock n' had above-median market capitalization on day d . Volume > 50%ile: An indicator variable that is one if stock n' had above-median trading volume on day d . Volatility > 50%ile: An indicator variable that is one if stock n' had above-median 1-minute-return volume on day d . Spread < 50%ile: An indicator variable that is one if stock n' had a below-median average bid-ask spread on day d . In Same Industry: An indicator variable that is one if stock n and stock n' both belonged to the same 3-digit SIC industry. (Reads) “There is little correlation between the LASSO’s choice of predictors and firm size, trading volume, return volatility, and bid-ask spread. And, the LASSO is less likely to select a stock from the same industry as a predictor.”*

The LASSO's Penalty Parameter

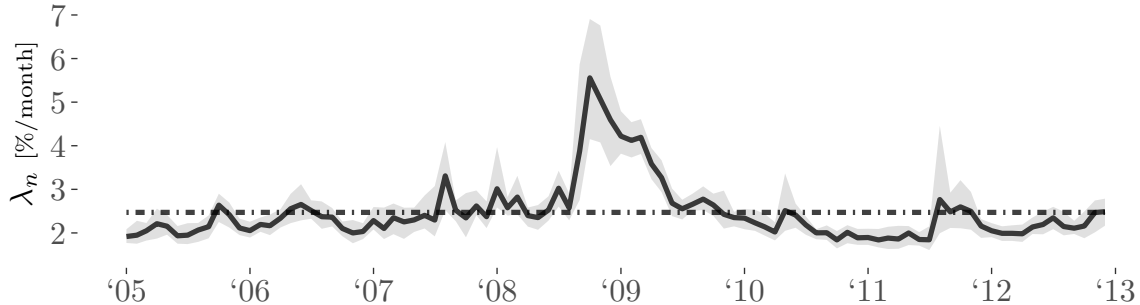


Figure 5: Monthly average (solid line) and 95% confidence interval (gray bands) for the LASSO's cross-validated penalty parameter, λ_n , in units of % per month. Dashed line denotes sample average over the course of the entire sample period, which is 2.5%. **(Sample)** Estimation results for a randomly selected subset of 250 stocks on each trading day from January 2005 to December 2012. **(Reads)** “The LASSO typically ignores all predictors weaker than $\lambda = 2.5\%$ per month when making its 1-minute-ahead return forecasts.”

erature. If they were to last an entire month, the predictors that we identify using the LASSO would result in outlandish returns that we do not see in practice. Put differently, existing steady long-lived predictors at weekly and monthly horizons are too weak at the 1-minute horizon for the LASSO to detect. This finding suggests that cross-sectional return predictability at the 1-minute horizon follows a different structure than cross-sectional return predictability at the 1-week and 1-month horizons.

4.2 Short-Lived

If the predictors identified by the LASSO would lead to unrealistic amounts of predictability if scaled up to the weekly or monthly horizons, then these predictors must not last that long. They must be short-lived. And, this is indeed the case. To show this, we compute the length of time that the LASSO typically uses each predictor for.

Econometric Approach. Suppose that stock n is one of the 250 randomly selected stocks for which we use the LASSO to make 1-minute-ahead return forecasts on day d . Now, consider the first minute that we generate a 1-minute-ahead return forecast, 10:04am. We collect all of the predictors used by the LASSO to make this return forecast for stock n . Then, in each subsequent minute, we compute the fraction of these predictors that are still being used by the LASSO to generate 1-minute-ahead return forecasts for stock n . If stock n' was used as a predictor in the LASSO's return

Predictor Duration

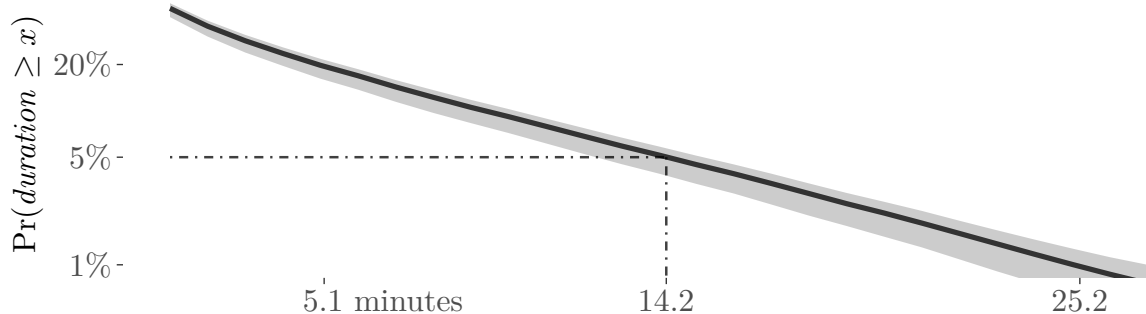


Figure 6: Average (solid line) and 95% confidence interval (gray bands) for the usage duration of the predictors chosen by the LASSO in units of minutes. y -axis is on a log scale and represents the probability that a predictor is used by the LASSO when making 1-minute-ahead return forecasts in more than x consecutive minutes. Tick marks on the x -axis represent the duration associated with a given quantile on the y -axis. **(Sample)** Estimation results for a randomly selected subset of 250 stocks on each trading day from January 2005 to December 2012. **(Reads)** “Less than 5% of predictors selected by the LASSO are used for more than 14.2 consecutive minutes.”

forecast in minute 10:04am but not in minute 10:05am, then we would say that this predictor had a duration of 1 minute. Whereas, if stock n' was used as a predictor in 10:04am as well as every subsequent minute up to but not including 10:14am, then we would say that this predictor had a duration of 10 minutes. We repeat this same calculation for each of the LASSO’s remaining 355 forecasts for stock n on that day. And, we do the same thing for every other stock on day d , and we do the same thing for every other trading day as well.

Empirical Results. Figure 6 shows the results of these calculations. We find that less than 5% of the predictors selected by the LASSO are used when making 1-minute-ahead return forecasts in more than 14.2 consecutive minutes. And, less than 1% of predictors are used by the LASSO for more than 25.2 minutes in a row. The short duration of the LASSO’s predictors helps explain the lack of correlation between these predictors and firm characteristics.

Some Intuition. Financial economists tend to study particular timescales. The field of market microstructure analyzes intraday events. People doing asset-pricing research consider phenomena at the daily, weekly, and monthly horizons. And, research on macrofinance topics answers questions about how markets evolve from one quarter or one year to the next.

Clearly, each of these horizon-specific subfields is studying different aspects of the same underlying return process. There is only one return process for the S&P 500. Its 1-minute returns in a month must sum up to its monthly return. So, it might seem odd that predictability at the 1-month horizon does not imply predictability at the 1-minute horizon. But, as illustrated in the parable of the blind men and the elephant (Holland and Quigley, 1959), the same underlying process can manifest itself in very different ways depending on the researcher’s perspective.

And, we are all familiar with processes that have different dynamics at different timescales in our everyday lives. Just think about how animals grow. A caterpillar hardly appears to change from one minute to the next, but it may well be a butterfly by this time next week. Or, think about how languages evolve. At this time last year, the word “awful” had pretty much the same meaning as it does today (very unpleasant; “an awful stench”). But, centuries ago the word had nearly the opposite meaning (inspiring wonder; “the awful majesty of God”).⁴

When we look at the kinds of predictors selected by the LASSO, we find something similar. Like caterpillars and butterflies, these predictors have little relation to the predictors at the weekly and monthly horizons documented in the academic literature. And, we can use the functional form of the LASSO’s optimization problem to shed light on how this can be the case. On one hand, in order to identify cross-sectional predictors fast enough, the LASSO has to ignore any source of predictability weaker than $\lambda = 2.5\%$ per month. On the other hand, the predictors identified by the LASSO typically last less than 15 minutes, making them imperceptible to monthly investors.

4.3 And Sparse

Finally, the LASSO should only increase out-of-sample fit if there are unexpected short-lived signals in the cross-section of returns that are also sparse. We find that, on average, the LASSO uses only 12.7 predictors (out of a possible $3 \cdot N \approx 6,000$) when making its 1-minute-ahead return forecasts for each stock. Figure 7 plots the average number of predictors, $S_{n,\lambda}$, used by the LASSO for each of its 1-minute-ahead return forecasts in a given month, and the gray bands report the 95% confidence interval.

⁴OED Online. Mar. 2017. Oxford Univ. Press. <https://goo.gl/DGkBE1>

Predictor Sparsity

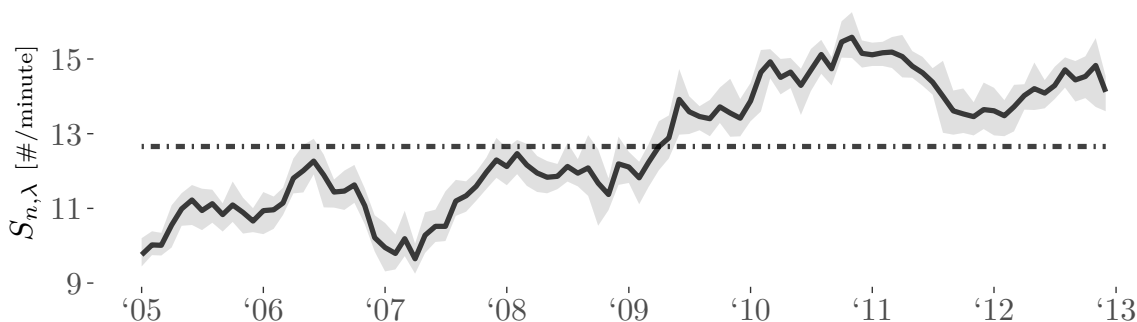


Figure 7: Average (solid line) and 95% confidence interval (gray bands) for the number of predictors, $S_{n,\lambda}$, used by the LASSO for each of its 1-minute-ahead return forecasts in a given month. Dashed line denotes sample average over the course of the entire sample period, which is 12.7 predictors. **(Sample)** Estimation results for a randomly selected subset of 250 stocks on each trading day from January 2005 to December 2012. **(Reads)** “On average, the LASSO uses only 12.7 predictors (out of a possible $3 \cdot N \approx 6,000!$) when making its 1-minute-ahead return forecasts for each stock.”

5 Economic Origins

The LASSO uses a purely statistical rule to identify candidate predictors that are too unexpected and short-lived to be pinned down by a researcher’s intuition alone. However, because it does not rely on a researcher’s intuition, the LASSO’s increase in out-of-sample fit could in principle have nothing to do with economic fundamentals. But, this is not what we find in the data. And, we conclude our analysis by showing that the LASSO tends to identify as predictors the lagged returns of stocks with news about firm fundamentals. In other words, the LASSO identifies predictors that are not easy to intuit but are still economically meaningful.

5.1 Ravenpack Data

To investigate the link between the LASSO’s choice of predictors and real-world events, we use data from Ravenpack.

Ravenpack has a partnership with Dow Jones, giving it access to the full Dow Jones news archive. This data consists of all Dow Jones Newswire and Wall Street Journal articles. The Dow Jones news archive have been used in many prior studies ([Kolasinski et al., 2013](#); [Shroff et al., 2013](#); [Dai et al., 2015](#)). Ravenpack data is timestamped

to the nearest millisecond, which allows us to examine the relative timing of news announcements and the LASSO’s choice of predictors.

Because we are particularly interested in the link between the LASSO’s choice of predictors and economically meaningful events, we look only at news articles in the Ravenpack data about a company’s revenues. In addition, we restrict our attention to articles that Ravenpack has labeled as “strongly relevant” for an NYSE-listed company. To make sure that we are not looking at stale news, we narrow our focus even further and only consider articles that Ravenpack labels as “news flashes” and gives the highest novelty score to. We also remove any news articles about companies which had another revenue-related news flash during the previous 5 days. Finally, we only consider news announcements that occur from 10:04am to 3:59pm. What we are left with is a collection of novel news flashes about firm revenues that occur during normal trading hours.

5.2 Event Study

Does the LASSO select a stock more often in the minutes around a news flash about the firm’s revenues? Yes.

Econometric Approach. Suppose that there is a new flash about the n ’th stock’s revenues in minute t . We define the event time, h , as the number of minutes since minute t . So, $h = 0$ denotes the minute that a news flash occurs, $h = 10$ denotes the 10th minute afterwards, and $h = -12$ denotes the 12th minute before. We then count the number of times that the LASSO uses stock n ’ as a predictor when creating its 1-minute-ahead return forecasts in each minute $h \in \{-30, \dots, -1, 0, 1, \dots, 30\}$. We use the LASSO to create 1-minute-ahead return forecasts for a randomly selected set of 250 stocks each minute, so the maximum number of times that the n ’th stock could be chosen by the LASSO in a given minute is 250.

Empirical Results. The LASSO uses a purely statistical rule to identify candidate predictors. So, the LASSO’s choice of predictors could in principle have nothing to do with economic fundamentals. If this were the case, then the number of times that the LASSO selected the n ’th stock would be constant with respect to event time, h . Figure 8 shows that this is not what is going on in the data. The x -axis counts minutes in event time. The dots denote the number of additional times that the LASSO selected the n ’th stock in minute h relative to minute $h = 0$. Red dots denote

The LASSO's Selection Rate Around News

	Dep. Variable: Times Selected [# / min]					
	Scheduled News			Unscheduled News		
$1_{\{-30 \leq h < 0\}}$	-0.115 (0.058)			0.113 (0.101)		
$1_{\{h=0\}}$	·			·		
$1_{\{30 \geq h > 0\}}$	-0.141 (0.054)			0.026 (0.091)		
$1_{\{-30 \leq h < -20\}}$	-0.139 (0.072)	-0.139 (0.073)		-0.029 (0.120)	-0.029 (0.120)	
$1_{\{-20 \leq h < -10\}}$	-0.124 (0.070)	-0.123 (0.070)		0.153 (0.124)	0.153 (0.124)	
$1_{\{-10 \leq h < 0\}}$	-0.086 (0.056)	-0.085 (0.056)		0.203 (0.094)	0.202 (0.094)	
$1_{\{h=0\}}$	·			·		
$1_{\{10 \geq h > 0\}}$	-0.156 (0.050)	-0.157 (0.050)		0.008 (0.086)	0.008 (0.086)	
$1_{\{20 \geq h > 10\}}$	-0.126 (0.063)	-0.127 (0.063)		0.001 (0.108)	0.001 (0.108)	
$1_{\{30 \geq h > 20\}}$	-0.142 (0.066)	-0.142 (0.066)		0.069 (0.109)	0.069 (0.109)	
Impact			0.201 (0.318)			-0.142 (0.876)
Sentiment			0.170 (0.209)			-1.090 (0.659)
Year FE	Y	Y	Y	Y	Y	Y

Table 8: *Change in the LASSO's selection rate around news flashes about a company's revenues. (Specification) Each column reports the results from a different regression. Point estimates have units of number of times selected per minute. Numbers in parentheses are standard errors clustered by year. Dots denote omitted estimates for the reference category, $h = 0$. (Variables) Dep. Variable: Number of times that the n' th stock was selected as a predictor by the LASSO in minute $(t + h)$. Impact: A zero (low) to one (high) variable summarizing news flash's impact on overall market volatility during the next several hours. Sentiment: A zero (very negative) to one (very positive) variable summarizing the sentiment of the text contained in a news flash. (Scheduled News) News flashes about scheduled events. (Unscheduled News) News flashes about unscheduled events. (Sample) 61-minute window around each novel news flash about an NYSE-listed stock's revenues which occurred during normal trading hours from January 2005 to December 2012. (Reads) "The LASSO uses stock n' to make 0.115 fewer 1-minute-ahead return forecasts in each of the 30 minutes leading up to a scheduled news flash, $1_{-30 \leq h < 0}$, than it does in the minute of the news flash itself, $h = 0$. By contrast, the LASSO uses stock n' to make 0.203 more 1-minute-ahead return forecasts in each of the 10 minutes leading up to an unscheduled news flash, $1_{-10 \leq h < 0}$."*

The LASSO's Selection Rate Around News

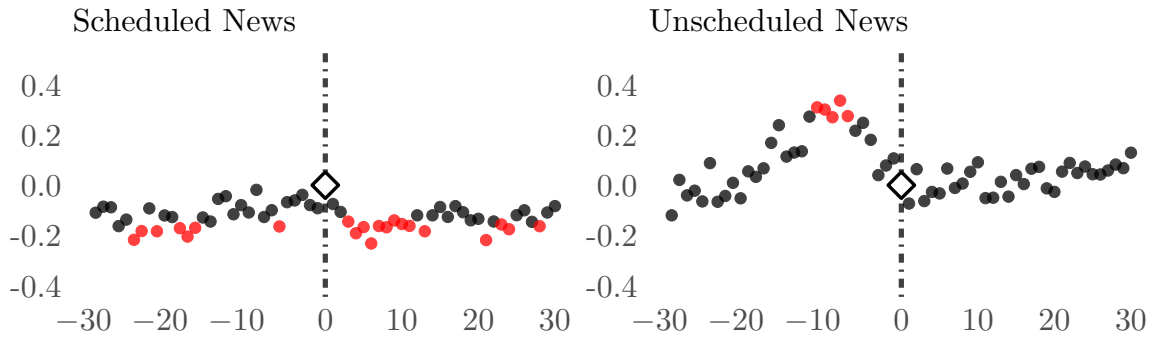


Figure 8: *x-axis: event time relative to news flash (in minutes). Vertical dashed line: minute of news flash about n' th stock's revenues, $h = 0$. y-axis: difference between the number of times that the LASSO selected the n' th stock as a predictor when making its 1-minute-ahead return forecast for minute h and the number of times it did so for minute $h = 0$. Each dot denotes the difference in the LASSO's selection rate in minute h relative to minute $h = 0$ with the large diamond denoting a difference of zero for $h = 0$ (a normalization). Red dots denote differences that are statistically significant at the 5% level using standard errors clustered by year. **(Scheduled News)** News flashes about scheduled events. **(Unscheduled News)** News flashes about unscheduled events. **(Sample)** 61-minute window around each novel news flash about an NYSE-listed stock's revenues which occurred during normal trading hours from January 2005 to December 2012. **(Reads)** "When there is a scheduled news flash about the n' th stock's revenues in minute t , the LASSO selects stock n' as a predictor slightly more often when making its 1-minute-ahead return forecasts for minute t —i.e., for $h = 0$. But, if the news flash is unscheduled, then the LASSO selects stock n' as a predictor much more often when making its 1-minute-ahead return forecasts in the 10 minutes prior to minute t ."*

differences that are statistically different from zero at the 5% level using standard errors that are clustered by stock. The left panel of Figure 8 reports the change in the LASSO's selection rate around scheduled news about firm revenues, revealing that the LASSO selects stock n' as a predictor slightly more often on the exact minute of the news flash. The right panel then reports the analogous results for unscheduled news. This panel shows that the LASSO selects stock n as a predictor when making its 1-minute-ahead return forecast much more often in the 10 minutes leading up to the unscheduled news flash.

Timing of Results. It might seem odd that the LASSO's selection rate increases in the 10 minutes *before* an unscheduled news flash. But, this lag makes sense upon further inspection. For example, think about 3M's announcement about cutting 1,800 jobs on December 8th, 2008. Or, consider St. Jude's announcement of the acquisition

of AGA medical on October 18th, 2010. The market did not know that these announcements were coming. And, as a result, there was a several minute lag between when these announcements were actually made and when journalists broadcast the resulting news flashes. By contrast, when journalists are expecting an announcement there should be no such gap. Consistent with this intuition, we find that the LASSO’s selection rate only increases for the exact minute of scheduled news flashes, not in the 10 minutes before.

Regression Analysis. Table 8 puts precise numbers on these results. Each column reports the results from a different regression. The dependent variable in all of these regressions is the number of times the n' th stock was selected as a predictor by the LASSO when making 1-minute-ahead return forecasts in a given minute. The first column shows that the LASSO uses stock n' to make 0.115 fewer 1-minute-ahead return forecasts in each of the 30 minutes leading up to a scheduled news flash than it does in the minute of the news flash and that the LASSO uses stock n' to make 0.141 fewer 1-minute-ahead return forecasts in each of the subsequent 30 minutes following a scheduled news flash. The fifth column of Table 8 reports that the LASSO uses stock n' to make 0.203 more 1-minute-ahead return forecasts in each of the 10 minutes leading up to an unscheduled news flash than it does in the minute of the news flash.

The third and sixth columns in Table 8 present two additional robustness check for our results. First, Ravenpack characterizes each news flash’s impact on overall market volatility over the course of the next several hours, and reports this information in a variable called “market impact” which takes on values from zero (no impact on aggregate volatility) to one (large impact on aggregate volatility). Consistent with our original hypothesis that the LASSO is identifying unexpected short-lived predictors, we find that the LASSO’s selection rate for stock n' is not related to the aggregate effect of a news announcement over the next several hours. Ravenpack also assigns each news flash a sentiment score ranging from zero (pessimistic) to one (optimistic). The sentiment of the news flash also appears unrelated to the LASSO’s selection rate in any statistically significant manner.

6 Conclusion

We apply the Least Absolute Shrinkage and Selection Operator (LASSO) rather than intuition to identify unexpected short-lived predictors, such as the lagged returns of

Family Dollar Corp. We find that using the LASSO increases both out-of-sample fit and forecast-implied Sharpe ratios, and we show that this increase in out-of-sample fit comes from identifying predictors that are unexpected, short-lived, and sparse. Finally, we document that these predictors are often the lagged returns of stocks with recent news about firm fundamentals. Thus, the LASSO identifies predictors that are not easy to intuit but still economically meaningful.

References

- Abadie, A. and M. Kasy (2017). The risk of machine learning. Working Paper.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6), 716–723.
- Arora, M. (2012). How secure is aes against brute force attacks? *EE Times* 5(7).
- Asness, C., T. Moskowitz, and L. Pedersen (2013). Value and momentum everywhere. *Journal of Finance* 68(3), 929–985.
- Banz, R. (1981). The relationship between the returns and the market values of common stocks. *Journal of Financial Economics* 9(1), 3–18.
- Boyd, S. and L. Vandenberghe (2004). *Convex Optimization*. Cambridge University Press.
- Brogaard, J., T. Hendershott, and R. Riordan (2014). High-frequency trading and price discovery. *Review of Financial Studies* 27(8), 2267–2306.
- Bryzgalova, S. (2017). Spurious factors in linear asset pricing models. Working Paper.
- Campbell, J. and S. Thompson (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21(4), 1509–1531.
- Candes, E. and Y. Plan (2009). Near-ideal model selection by ℓ_1 minimization. *Annals of Statistics* 37(5a), 2145–2177.
- Carhart, M. (1997). On persistence in mutual fund performance. *Journal of Finance* 52(1), 57–82.
- Chinco, A. (2015). Feature-selection risk. Working Paper.
- Chordia, T. and B. Swaminathan (2000). Trading volume and cross-autocorrelations in stock returns. *Journal of Finance* 55(2), 913–935.
- Cohen, L. and A. Frazzini (2008). Economic links and predictable returns. *Journal of Finance* 63(4), 1977–2011.
- Dai, L., J. Parwada, and B. Zhang (2015). The governance role of the media through news dissemination: evidence from insider trading. *Journal of Accounting Research* 53(2), 331–366.
- Dangl, T. and M. Halling (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics* 106(1), 157–181.

- DeMiguel, V., L. Garlappi, F. Nogales, and R. Uppal (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science* 55(5), 798–812.
- DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies* 22(5), 1915–1953.
- DeMiguel, V., F. Nogales, and R. Uppal (2014). Stock return serial dependence and out-of-sample portfolio performance. *Review of Financial Studies* 27(4), 1031–1073.
- Du, S. and H. Zhu (2016). What is the optimal trading frequency in financial markets. *Review of Financial Studies*, Forthcoming.
- Efron, B. and T. Hastie (2016). *Computer Age Statistical Inference* (1 ed.). Cambridge University Press.
- Fama, E. and K. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics* 33(1), 3–56.
- Feng, G., S. Giglio, and D. Xiu (2017). Taming the factor zoo. Working Paper.
- Ferson, W. and C. Harvey (1991). The variation of economic risk premiums. *Journal of Political Economy* 99(2), 385–415.
- Foster, D. and E. George (1994). The risk inflation criterion for multiple regression. *Annals of Statistics* 22(4), 1947–1975.
- Foucault, T., J. Hombert, and I. Rosu (2016). News trading and speed. *Journal of Finance* 71(1), 335–382.
- Frazzini, A., R. Israel, and T. Moskowitz (2015). Trading costs of asset pricing anomalies. Working Paper.
- Freyberger, J., A. Neuhierl, and M. Weber (2017). Dissecting characteristics non-parametrically. Working Paper.
- Friedman, J., T. Hastie, and R. Tibshirani (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33(1), 1–22.
- Gabaix, X. (2014). A sparsity-based model of bounded rationality. *Quarterly Journal of Economics* 129(4), 1661–1710.
- Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica* 74(6), 1545–1578.

- Harvey, C., Y. Liu, and H. Zhu (2016). . . . and the Cross-Section of expected returns. *Review of Financial Studies* 29(1), 5–68.
- Hasbrouck, J. (2009). Trading costs and returns for u.s. equities: Estimating effective costs from daily data. *Journal of Finance* 64(3), 1445–1477.
- Hastie, T., R. Tibshirani, and J. Friedman (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (1 ed.). Springer Science & Business Media.
- Hoffman, P. (2014). A dynamic limit order market with fast and slow traders. *Journal of Financial Economics* 113(1), 156–169.
- Holland, J. and L. Quigley (1959). *The blind men and the elephant*. Charles Scribner’s Sons.
- Hou, K. (2007). Industry information diffusion and the lead-lag effect in stock returns. *Review of Financial Studies* 20(4), 1113–1138.
- Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48(1), 65–91.
- Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2014). Time-varying fund manager skill. *Journal of Finance* 69(4), 1455–1484.
- Kolasinski, A., A. Reed, and M. Ruggenberg (2013). A multiple lender approach to understanding supply and search in the equity lending market. *Journal of Finance* 68(2), 559–595.
- Korajczyk, R. and R. Sadka (2004). Are momentum profits robust to trading costs? *Journal of Finance* 59(3), 1039–1082.
- Kozak, S., S. Nagel, and S. Santosh (2017). Shrinking the cross-section. Working Paper.
- Linnainmaa, J. and M. Roberts (2016). The history of the cross section of stock returns. Working Paper.
- Lo, A. and C. MacKinlay (1990). When are contrarian profits due to stock-market overreaction? *Review of Financial Studies* 3(2), 175–205.
- Manela, A. and A. Moreira (2017). News-implied volatility and disaster concerns. *Journal of Financial Economics* 123(1), 137–162.
- McLean, D. and J. Pontiff (2016). Does academic research destroy stock return predictability? *Journal of Finance* 71(1), 5–32.

- Meinshausen, N. and B. Yu (2009). LASSO-type recovery of sparse representations for high-dimensional data. *Annals of Statistics* 37(1), 246–270.
- Natarajan, B. (1995). Sparse approximate solutions to linear systems. *SIAM Journal of Computations* 24(2), 227–234.
- Novy-Marx, R. (2016). Backtesting strategies based on multiple signals. Working Paper.
- Novy-Marx, R. and M. Velikov (2015). A taxonomy of anomalies and their trading costs. *Review of Financial Studies* 29(1), 104–147.
- Patton, A. and B. Weller (2017). What you see is not what you get: The costs of trading market anomalies. Working Paper.
- Pedersen, L. (2015). *Efficiently Inefficient: How Smart Money Invests and Market Prices Are Determined* (1 ed.). Princeton University Press.
- Rapach, D., J. Strauss, and G. Zhou (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, 821–862.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6(2), 461–464.
- Shroff, N., R. Verdi, and G. Yu (2013). Information environment and the investment decisions of multinational corporations. *Accounting Review* 89(2), 759–790.
- Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B* 58(1), 267–288.
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 1418–1429.
- Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B* 67(2), 301–320.

A Numerical Simulations

In this appendix, we use the LASSO to make 1-minute-ahead return forecasts in simulated data. The goal is to show that, if there are sparse signals in the cross-section of returns, then the LASSO increases out-of-sample fit. And, if there are not (either no signals or too many signals), then it does not.

Data-Generating Process. We run 1,000 simulations. For each simulation, we generate 1-minute returns for 100 stocks over the course of $T = 1301$ minutes using the following data-generating process:

$$r_{n,t} = 0 + \sum_{n'=1}^{100} \beta_{(n,n'),t}^* \cdot r_{n',t-1} + 0.001 \cdot \varepsilon_{n,t}^* \quad (19)$$

In the equation above, $r_{n,t}$ is the return of the n th stock in minute t , $\beta_{(n,n'),t}^*$ is the predictive power of the n' th stock's lagged return in minute $(t-1)$ when forecasting the n th stock's return in minute t , and $\varepsilon_{n,t}^* \stackrel{\text{iid}}{\sim} \text{N}(0, 1)$ is a noise term. For computational convenience, we only consider a single lag.

In our main analysis, each stock's return in minute t is governed by its exposure to the lagged returns of a collection of 5 stocks in minute $(t-1)$. We use \mathcal{S}_t to denote this active set. For all 95 stocks $n' \notin \mathcal{S}_t$, $\beta_{(n,n'),t}^* = 0$. For all $n' \in \mathcal{S}_t$, we choose the values of $\beta_{(n,n'),t}^*$ as follows:

$t = 1$: Prior to the start of trading, we initialize the model by randomly selecting a set of 5 stocks to serve as \mathcal{S}_1 . We determine the exposure of every stock $n \in \{1, \dots, 100\}$ to the lagged returns of the 5 stocks $n' \in \mathcal{S}_1$ using the rule:

$$\beta_{(n,n'),1}^* \stackrel{\text{iid}}{\sim} \begin{cases} +0.19 & \text{w/ prob. } 50\% \\ -0.19 & \text{w/ prob. } 50\% \end{cases} \quad (20)$$

$t > 1$: Prior to each subsequent minute, we draw 5 random variables $z_{n',t} \stackrel{\text{iid}}{\sim} \text{Unif}[0, 1]$ —i.e., one for each $n' \in \mathcal{S}_{t-1}$. If a particular $z_{n',t} > 0.01$, then predictor $n' \in \mathcal{S}_{t-1}$ remains in \mathcal{S}_t . And, all 100 values of $\beta_{(n,n'),t}^*$ associated with this predictor remain the same. But, if $z_{n',t} \leq 0.01$, then predictor $n' \in \mathcal{S}_{t-1}$ is replaced at random with one of the remaining 95 stocks. And, we redraw each of the 100 stocks' exposures to this new predictor exactly as described in Equation (20).

Note that, once a predictor gets added to \mathcal{S}_t , it remains in the active set for $(1 - 0.01)/0.01 = 99$ minutes on average. There are 100 stocks. So, the collection of 5

Out-of-Sample Fit, Simulated Data

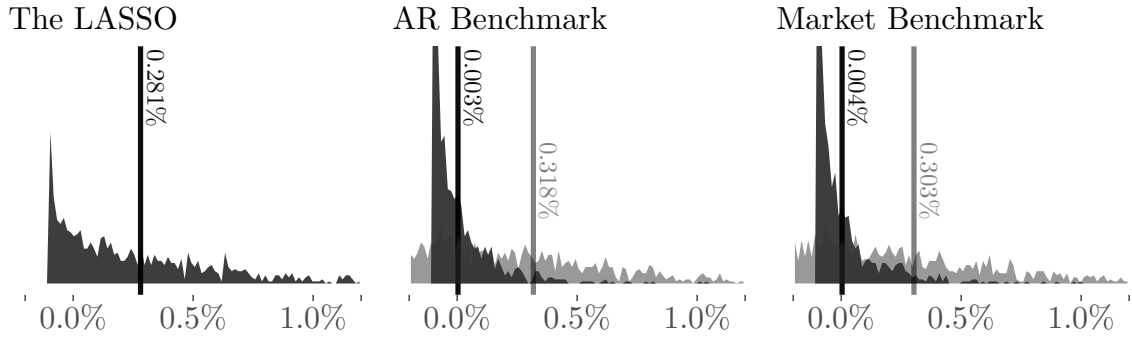


Figure 9: *Out-of-sample fit from predictive regressions using simulated data generated by Equation (19). x-axis: adjusted R^2 from predictive regression in percent. (The LASSO) Distribution of adjusted R^2 s in predictive regressions involving only the LASSO’s 1-minute-ahead return forecast. Vertical line denotes the average adjusted R^2 from these regressions. (AR Benchmark) Dark-gray shading denotes the distribution of adjusted R^2 s in predictive regressions involving only the AR(1) benchmark’s 1-minute-ahead return forecast. Light-gray shading denotes the distribution of adjusted R^2 s in predictive regressions involving the 1-minute-ahead return forecasts from both the LASSO and the AR(1) benchmark. Vertical lines denote the average adjusted R^2 s from each specification. (Market Benchmark) Dark-gray shading denotes the distribution of adjusted R^2 s in predictive regressions involving only the market benchmark’s 1-minute-ahead return forecast. Light-gray shading denotes the distribution of adjusted R^2 s in predictive regressions involving the 1-minute-ahead return forecasts from both the LASSO and the market benchmark. Vertical lines denote the average adjusted R^2 s from each specification. (Reads) “Including the LASSO’s 1-minute-ahead return forecast boosts the out-of-sample fit in simulated data.”*

predictors in the active set is too short-lived to be estimated with an OLS regression.

Estimation and Forecast Timing. We use the LASSO to make 1-minute-ahead return forecasts for stock $n = 1$ in periods $t = 302, \dots, 1301$ exactly as we did in the main text, only using a $L = 50$ -minute estimation window. We burn the first 250 periods to make sure that the initial conditions are not affecting our results. These procedure produces a total of 1000 return forecasts in each simulation. We also fit two different OLS specifications to the same data. The first specification is an AR(1) model which only uses a stock’s own lagged return as a predictor. The second specification is a market model which uses lagged values of the equally-weighted market as a predictor.

Out-of-Sample Fit. The dark-gray shading in the left panel of Figure 9 shows that the average out-of-sample fit of the LASSO’s 1-minute-ahead return forecast is

Oracle Out-of-Sample Fit, Simulated Data

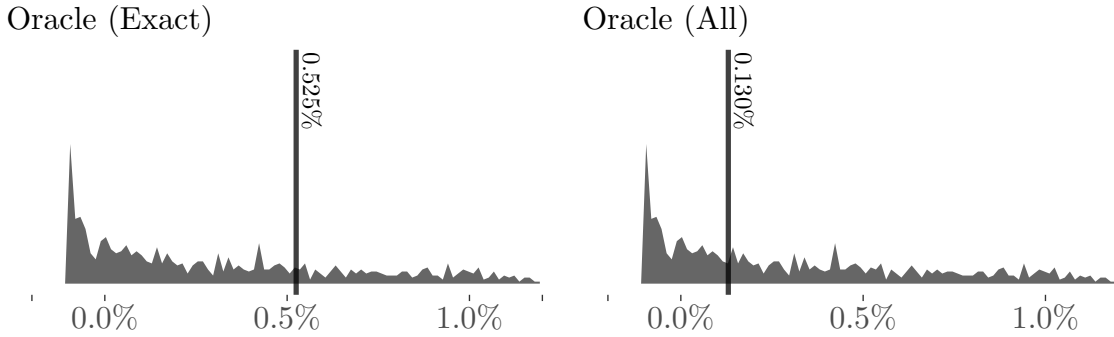


Figure 10: *Out-of-sample fit from predictive regressions using simulated data generated by Equation (19). x-axis: adjusted R^2 from predictive regression in percent. (Oracle, Exact) Distribution of adjusted R^2 s in predictive regressions involving an oracle specification using the exact active set. Vertical line denotes the average adjusted R^2 from these regressions. (Oracle, All) Distribution of adjusted R^2 s in predictive regressions involving an oracle specification using every predictor that ever belonged to the active set during the estimation window. Vertical line denotes the average adjusted R^2 from these regressions. (Reads) “The maximum out-of-sample fit that you could achieve—i.e., if you knew exactly which 5 predictors to include in your OLS regression—is 0.525%.”*

0.281%. Just like in Section 3.2, we measure the LASSO’s out-of-sample fit using the adjusted R^2 statistic. This explains why the out-of-sample fit can sometimes be negative. By contrast, the dark-gray shading in middle panel shows that the average out-of-sample fit of the AR(1) benchmark is only 0.003%, while the dark-gray shading in the right panel shows that the average out-of-sample fit of the market benchmark is only 0.004%. The light-gray shading in each of these panels displays the out-of-sample fit of using both the LASSO and a benchmark model to predict future returns. In both cases, the LASSO significantly increases in the out-of-sample fit, boosting the average adjusted R^2 to more than 0.300%.

Oracle Regression. Is an adjusted $R^2 \approx 0.300\%$ big or small? To get a sense of how much out-of-sample fit is possible given the data-generating process in Equation (19), we make 1-minute-ahead return forecasts using two different versions of an oracle estimator. First, we estimate an OLS regression using the 5 true predictors governing the cross-section of returns in minute t , $n' \in \mathcal{S}_t$. The left panel of Figure 10 shows that, even if you knew the correct 5 predictors to use each minute, the best adjusted R^2 you could hope for would be 0.525%. So, the LASSO is capturing a little more than half of the possible variation in returns.

Number of Predictors, Simulated Data

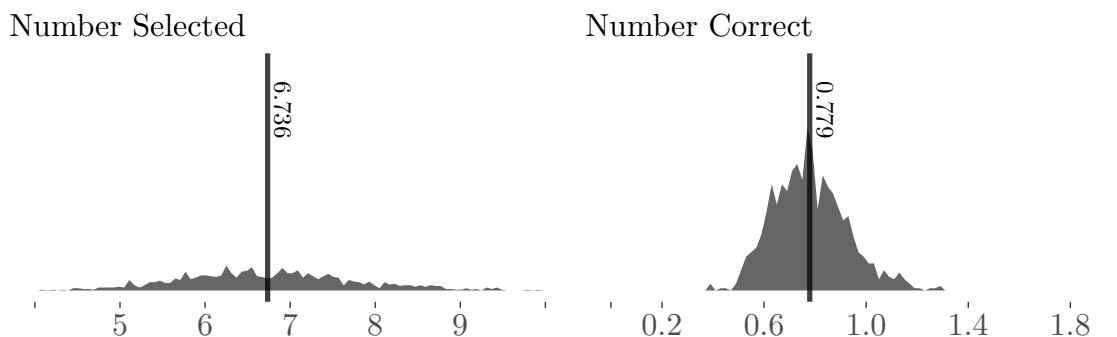


Figure 11: *The LASSO’s choice of predictors. (Number Selected) Number of predictors selected by the LASSO each minute when making its 1-minute-ahead return forecasts using the data simulated from Equation (19). Vertical line denotes the sample average across all simulations. (Number Correct) Number of predictors selected by the LASSO that belong to the true active set, \mathcal{S}_t . Vertical line denotes the sample average across all simulations. (Reads) “The LASSO selects a predictor in the active set roughly 80% of the time.”*

This first oracle specific is labeled as “Oracle (Exact)” in Figure 10. We use this name because this specification knows exactly which 5 predictors to use when making return forecasts each minute during the 50-minute estimation window. But, during any given estimation window, the active set of 5 predictors could change. Even though there are only 5 stocks in \mathcal{S}_t at any one point in time, there might be a total of 6 or 7 stocks that were each included in \mathcal{S}_t at some point in time during an estimation window. To see how accounting for these additional predictors would affect our results, we also estimate an oracle specification which uses all stocks which belonged to \mathcal{S}_t at some point during the estimation window. The right panel of Figure 10, which we label as “Oracle (All)” shows that including every stock that ever belonged to the active set drops the out-of-sample fit to 0.130%.

This second version of the oracle estimator actually fares worse than the LASSO’s because only 5 predictors belong to the active set, \mathcal{S}_t , that is used to generate the returns for minute t . Any additional predictors that were included in some $\{\mathcal{S}_{t-h}\}_{0 < h \leq 50}$ but not in \mathcal{S}_t will not add any out-of-sample predictive power. They are spurious predictors. And, the LASSO’s penalty function mitigates the problem posed by such spurious predictors.

Number of Predictors. The left panel of Figure 11 shows that the LASSO typically selects 6.736 predictors when making its 1-minute-ahead return forecasts each minute.

Out-of-Sample Fit, Simulated Data With No Predictors

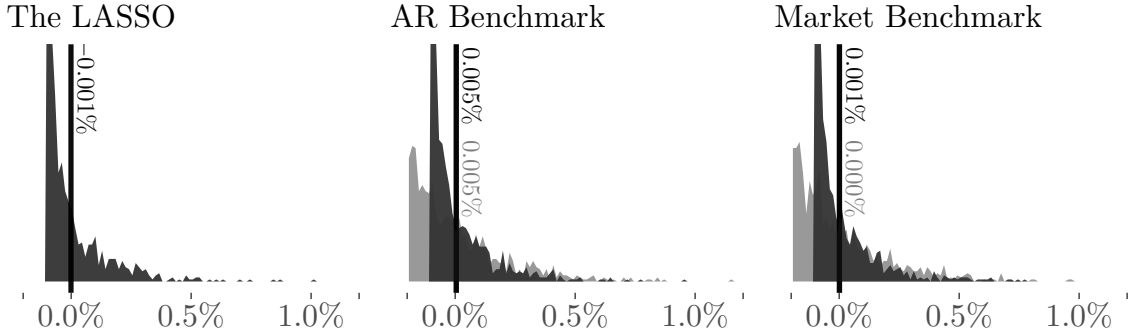


Figure 12: *Out-of-sample fit from predictive regressions using simulated data generated by Equation (21), which contains no predictors for the LASSO to find. x-axis: adjusted R^2 from predictive regression in percent. (The LASSO) Distribution of adjusted R^2 s in predictive regressions involving only the LASSO’s 1-minute-ahead return forecast. Vertical line denotes the average adjusted R^2 from these regressions. (AR Benchmark) Dark-gray shading denotes the distribution of adjusted R^2 s in predictive regressions involving only the AR(1) benchmark’s 1-minute-ahead return forecast. Light-gray shading denotes the distribution of adjusted R^2 s in predictive regressions involving the 1-minute-ahead return forecasts from both the LASSO and the AR(1) benchmark. Vertical lines denote the average adjusted R^2 s from each specification. (Market Benchmark) Dark-gray shading denotes the distribution of adjusted R^2 s in predictive regressions involving only the market benchmark’s 1-minute-ahead return forecast. Light-gray shading denotes the distribution of adjusted R^2 s in predictive regressions involving the 1-minute-ahead return forecasts from both the LASSO and the market benchmark. Vertical lines denote the average adjusted R^2 s from each specification. (Reads) “The LASSO does not increase out-of-sample fit when there are no predictors.”*

The right panel of Figure 11 shows that, on average, only 0.779 of these predictors belongs to the true active set, \mathcal{S}_t , in any given minute. It might seem surprising that the LASSO can achieve more than 1/2 of the oracle’s out-of-sample fit while correctly identifying less than 1/5 of the active set. But, this is another consequence of the LASSO’s penalty function, which shrinks weak predictors toward zero. The fact that the LASSO has such a high out-of-sample fit means that it is substantially shrinking the $6.736 - 0.779 = 5.957$ spurious predictors it selects each minute.

Placebo Tests. Finally, we conclude this appendix by looking at two alternative simulations where the LASSO should not increase out-of-sample fit. In the first setting, there are no predictors—i.e., we simulate the returns for the 100 stocks using the data-generating process below:

$$r_{n,t} = 0.125 \times \{0.001 \cdot \varepsilon_{n,t}^*\} \quad (21)$$

The factor of 0.125 is chosen so that the resulting time series have the same standard deviation as in Equation (19). Figure 12 confirms that, when there are no predictors for the LASSO to find, using the LASSO does not increase out-of-sample fit.

Then, in the second setting, we look at the other extreme where the predictors are dense—i.e., we simulate the returns for the 100 stocks using the data-generating process in Equation (19) but with 75 rather than 5 stocks in the active set. Again, to keep the standard deviation of returns constant, we choose $\beta_{(n,n'),t}^*$ using the rule:

$$\beta_{(n,n'),t}^* = \frac{5}{75} \times \begin{cases} +0.19 & \text{w/ prob. 50\%} \\ -0.19 & \text{w/ prob. 50\%} \end{cases} \quad (22)$$

Figure 13 confirms that, when there are more predictors in the active set, $|\mathcal{S}_t| = 75$, than minutes in the estimation window, $L = 50$, the LASSO does not increase out-of-sample fit. Thus, the LASSO really is only identifying sparse signals in the cross-section of returns.

B Significance Test

We use the approach from [Giacomini and White \(2006\)](#) to assess statistical significance using a Wald-type test outlined in Theorem 1. The results in [Giacomini and White \(2006, Thm. 1\)](#) are framed in terms of loss in forecasting power, $\Delta L_{n,t+1}$, and a test function, h_t . Whereas, we frame our results in terms of changes in R^2 . Here is the connection.

First, let $e_{n,t+1}^{\text{Bmk}}$ denote the prediction error when using only the 1-minute-ahead return forecast from a benchmark model, and let $e_{n,t+1}^{\text{Both}}$ denote the prediction error when using the 1-minute-ahead return forecasts from both the benchmark model and the LASSO. And, let the loss in forecasting power be the increase in squared prediction error from using only the benchmark model:

$$\Delta L_{n,t+1} \stackrel{\text{def}}{=} (e_{n,t+1}^{\text{Bmk}})^2 - (e_{n,t+1}^{\text{Both}})^2 \quad (23)$$

If using the LASSO in addition to the benchmark model does not increase out-of-sample fit, then:

$$\text{E}[\Delta L_{n,t+1}] = 0 \quad (24)$$

This is the null hypothesis. The alternative hypothesis is that $\text{E}[\Delta L_{n,t+1}] > 0$.

Next, notice that the expected loss in forecasting power divided by the realized

Out-of-Sample Fit, Simulated Data With Dense Predictors

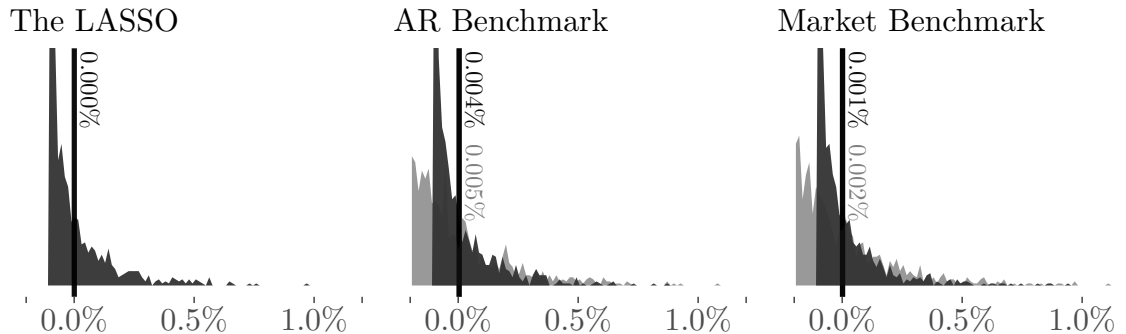


Figure 13: *Out-of-sample fit from predictive regressions using simulated data generated by Equation (19) with an active set containing 75 rather than 5 stocks. x-axis: adjusted R^2 from predictive regression in percent. (The LASSO) Distribution of adjusted R^2 s in predictive regressions involving only the LASSO’s 1-minute-ahead return forecast. Vertical line denotes the average adjusted R^2 from these regressions. (AR Benchmark) Dark-gray shading denotes the distribution of adjusted R^2 s in predictive regressions involving only the AR(1) benchmark’s 1-minute-ahead return forecast. Light-gray shading denotes the distribution of adjusted R^2 s in predictive regressions involving the 1-minute-ahead return forecasts from both the LASSO and the AR(1) benchmark. Vertical lines denote the average adjusted R^2 s from each specification. (Market Benchmark) Dark-gray shading denotes the distribution of adjusted R^2 s in predictive regressions involving only the market benchmark’s 1-minute-ahead return forecast. Light-gray shading denotes the distribution of adjusted R^2 s in predictive regressions involving the 1-minute-ahead return forecasts from both the LASSO and the market benchmark. (Reads) “The LASSO does not increase out-of-sample fit when there are more predictors than minutes in the estimation window.”*

variation in returns can be written as the change in R^2 :

$$\frac{\overline{\mathbb{E}}[\Delta L_{n,t+1}]}{\overline{\text{Var}}[r_{n,t+1}]} = \frac{\overline{\mathbb{E}}[(e_{n,t+1}^{\text{Bmk}})^2]}{\overline{\text{Var}}[r_{n,t+1}]} - \frac{\overline{\mathbb{E}}[(e_{n,t+1}^{\text{Both}})^2]}{\overline{\text{Var}}[r_{n,t+1}]} \quad (25a)$$

$$= -1 + \frac{\overline{\mathbb{E}}[(e_{n,t+1}^{\text{Bmk}})^2]}{\overline{\text{Var}}[r_{n,t+1}]} + 1 - \frac{\overline{\mathbb{E}}[(e_{n,t+1}^{\text{Both}})^2]}{\overline{\text{Var}}[r_{n,t+1}]} \quad (25b)$$

$$= \left(1 - \frac{\overline{\mathbb{E}}[(e_{n,t+1}^{\text{Both}})^2]}{\overline{\text{Var}}[r_{n,t+1}]}\right) - \left(1 - \frac{\overline{\mathbb{E}}[(e_{n,t+1}^{\text{Bmk}})^2]}{\overline{\text{Var}}[r_{n,t+1}]}\right) \quad (25c)$$

$$= \bar{R}_n^{2,\text{Both}} - \bar{R}_n^{2,\text{Bmk}} \quad (25d)$$

$$= \Delta \bar{R}_n^2 \quad (25e)$$

Finally, assume that $h_t = 1$. We know that $\Delta \bar{R}_n^2 \rightarrow \chi_1^2$, so [Giacomini and White \(2006, Thm. 1\)](#) implies that we can use a standard t -test to evaluate the statistical significance of changes in R^2 . The authors highlight this observation in [Comment 7](#).